Targeting 24 bp within Telomere Repeat Sequences with Tandem Tetramer Pyrrole-Imidazole Polyamide Probes

AUTHOR(S):

Kawamoto, Yusuke; Sasaki, Asuka; Chandran, Anandhakumar; Hashiya, Kaori; Ide, Satoru; Bando, Toshikazu; Maeshima, Kazuhiro; Sugiyama, Hiroshi

CITATION:

Kawamoto, Yusuke ...[et al]. Targeting 24 bp within Telomere Repeat Sequences with Tandem Tetramer Pyrrole-Imidazole Polyamide Probes. Journal of the American Chemical Society 2016, 138(42): 14100-14107

ISSUE DATE:

2016-10-26

URL:

http://hdl.handle.net/2433/230915

Targeting 24 bp within Telomere Repeat Sequences with Tandem Tetramer Pyrrole–Imidazole Polyamide Probes

Yusuke Kawamoto[†], Asuka Sasaki[‡], Anandhakumar Chandran[†], Kaori Hashiya[†] Satoru Ide[‡], Toshikazu Bando[*,†], Kazuhiro Maeshima[*,‡], and Hiroshi Sugiyama[*,†,§]

[†]Department of Chemistry, Graduate School of Science, Kyoto University, Sakyo, Kyoto 606–8502, Japan.

[‡]Biological Macromolecules Laboratory, Structural Biology Center, National Institute of Genetics, and Department of Genetics, School of Life Science, Graduate University for Advanced Studies (Sokendai), Mishima, Shizuoka 411–8540, Japan.

[§]Institute for Integrated Cell–Material Science (WPI–iCeMS), Kyoto University, Sakyo, Kyoto 606–8501, Japan.

ABSTRACT: Synthetic molecules that bind sequence-specifically to DNA have been developed for varied biological applications, including anticancer activity, regulation of gene expression, and visualization of specific genomic regions. Increasing the number of base pairs targeted by synthetic molecules strengthens their sequence specificity. Our group has been working on the development of pyrrole–imidazole polyamides that bind to the minor groove of DNA in a sequence-specific manner without causing denaturation. Recently, we reported a simple synthetic method of fluorescent tandem dimer polyamide probes composed of two hairpin moieties with a linking hinge, which bound to 12 bp in human telomeric repeats (5′–$(TTAGGG)_n$–3′) and could be used to specifically visualize telomeres in chemically fixed cells under mild conditions. We also performed structural optimization and extension of the target base pairs to allow more specific staining of telomeres. In the present study, we synthesized tandem tetramer polyamides composed of four hairpin moieties, targeting 24 bp in telomeric repeats, the longest reported binding site for synthetic, non-nucleic-acid-based, sequence-specific DNA-binding molecules. The novel tandem tetramers bound with a nanomolar dissociation constant to 24 bp sequences made up of four telomeric repeats. Fluorescently labeled tandem tetramer polyamide probes could visualize human telomeres in chemically fixed cells with lower background signals than polyamide probes reported previously, suggesting that they had higher specificity for telomeres. Furthermore, high-throughput sequencing of human genomic DNA pulled down by the biotin-labeled tandem tetramer polyamide probe confirmed its effective binding to telomeric repeats in the complex chromatinized genome.

Introduction

Sequence-specific DNA binding molecules have potential as DNA-based therapeutics and diagnostics. Various kinds of programmable, sequence-specific DNA-binding molecules have been developed, including the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated caspase 9 (Cas9) systems,[1] protein-based zinc fingers,[2] transcription activator-like effector (TALE),[3] and nucleic-acid-based triplex-forming oligonucleotides (TFO).[4] Chemically modified nucleic acid analogues such as peptide nucleic acids (PNA) can also function as TFOs.[5] Other synthetic, non-nucleic-acid-based sequence-specific DNA-binding molecules have been reported, including amidine-benzimidazole-phenyl (ABP)-based minor groove binders,[6] DNA-binding fragments of transcription factors,[7] threading polyintercalators[8] and minor groove-binding pyrrole–imidazole (Py–Im) polyamides.[9] Our group has been working on pyrrole–imidazole (Py–Im) polyamides first described by Dervan and coworkers. Py–Im polyamides are composed of *N*-methylpyrrole, *N*-methylimidazole, and aliphatic amino acids connected by amide bonds, and they bind sequence-specifically to the minor groove of double-stranded B-DNA by recognizing Watson–Crick base pairs. The binding rule is that antiparallel arrangement of Im/Py can bind to G•C base pairs, whereas Py/Py can bind to both A•T and T•A base pairs.[9f,g] Referred to as aliphatic amino acid residues, the C-terminal β-alanine tail and γ-aminobutyric acid or 2,4-diaminobutyric acid (Dab) turn moiety, called a γ-turn, also binds to A•T and T•A base pairs.[10] The γ-turn connects two aromatic peptides to form hairpin[11] and cyclic[12] structures, and the C-terminal β-alanine tail increases sequence specificity and binding affinity.[13a] In contrast, internal β-alanine can form β/β, β/Py and β/Im pairs as a substitution for Py rings and can relax the curvature of the polyamides to allow their efficient binding.[13b] Py–Im polyamides can be synthesized easily by machine-assisted solid-phase peptide

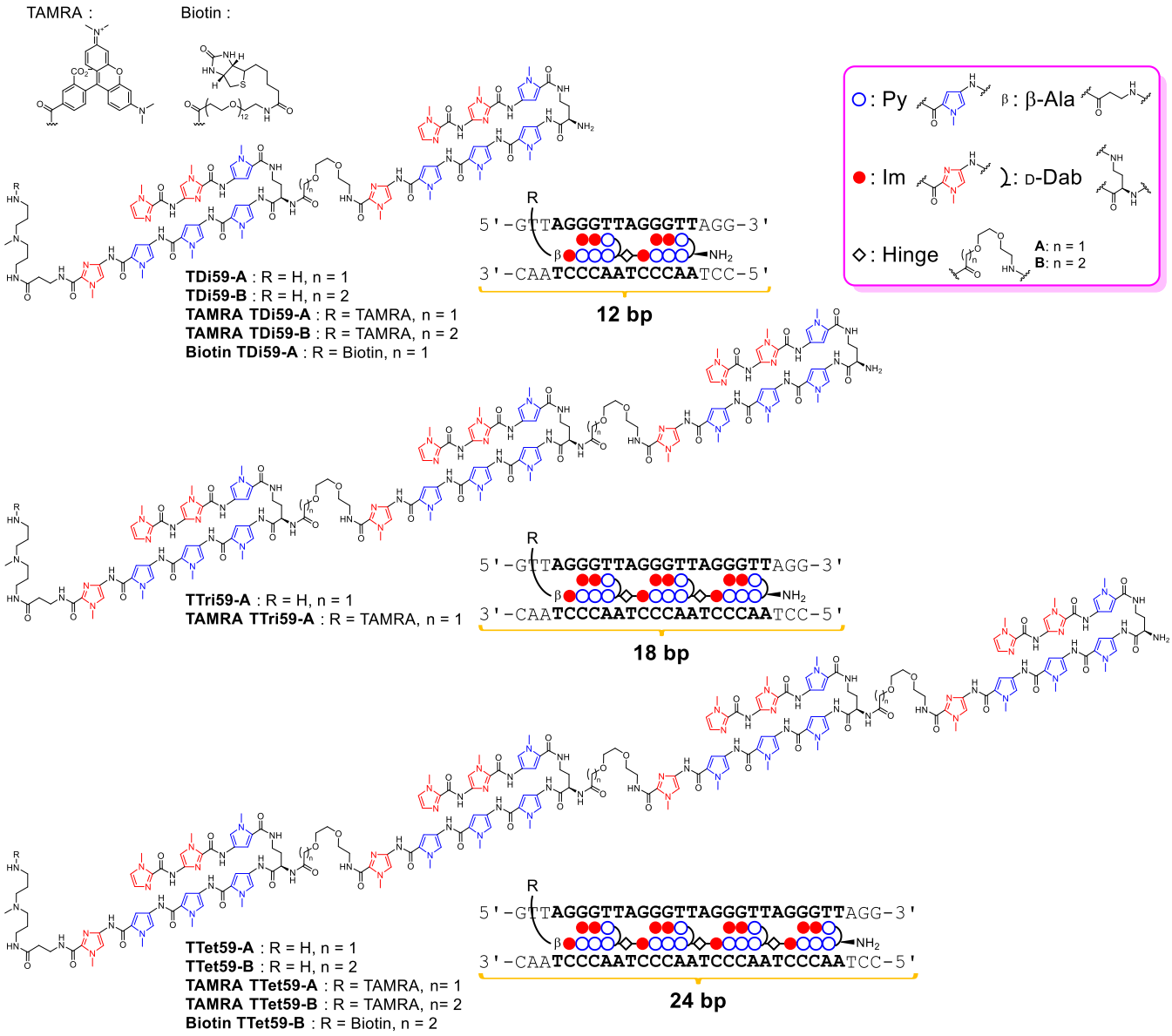


Figure 1. Chemical structures and schematic representations of tandem dimer **TDi59–A** and **–B**, **TAMRA TDi59–A** and **–B** and **Biotin TDi59–A**, tandem trimer **TTri59–A** and **TAMRA TTri59–A**, and tandem tetramer **TTet59–A** and **–B** and **TAMRA TTet59–A** and **–B** and **Biotin TTet59–B** targeting for 12 bp, 18 bp, and 24 bp in human telomere sequence, respectively.

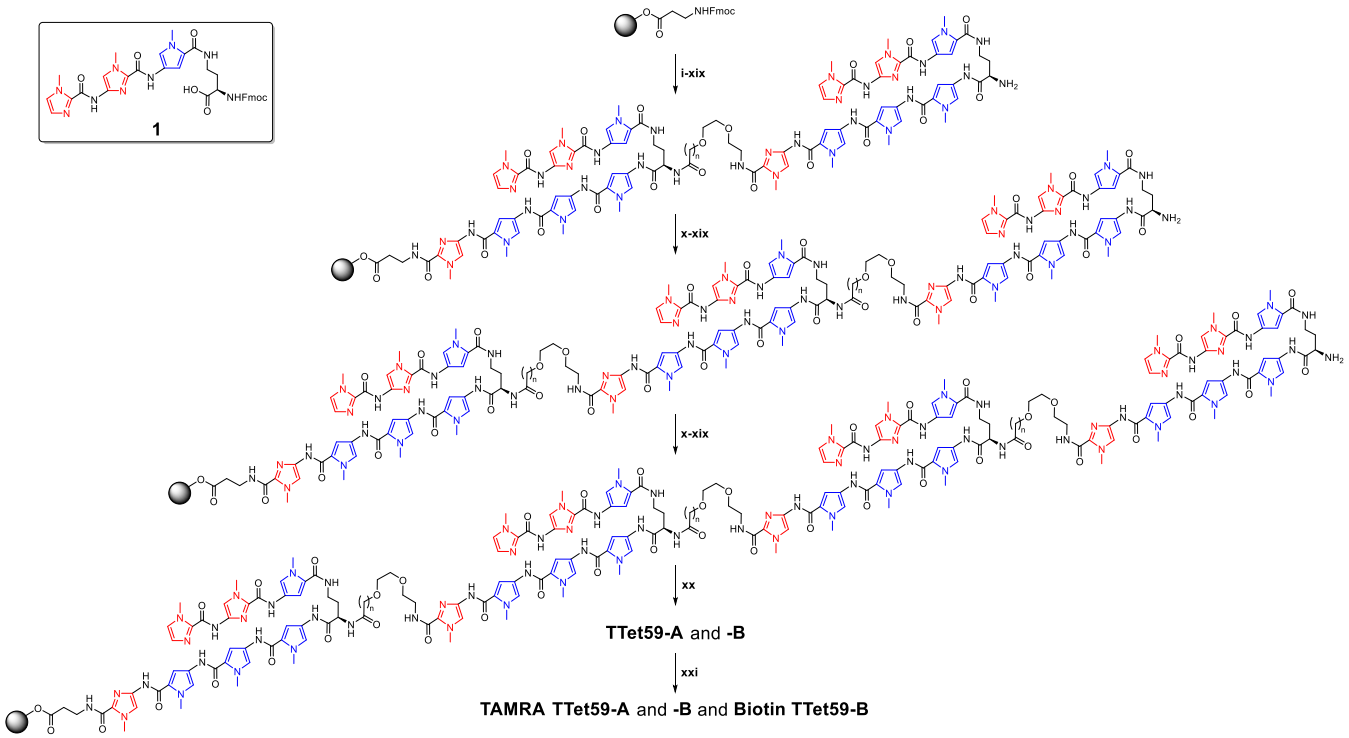
synthesis (SPPS)[12d,14] and have been used for many purposes; for example, visualization of specific sequences,[15] pull-down and identification of binding regions,[16] inhibition of gene expression,[17] transcriptional activation,[18,19] and detection of specific sequences on a solid substrate.[20] To target a single site in the whole human genome, which comprises about 3 billion base pairs, a sequence of ≥ 16 bp must be recognized by DNA-binding molecules; therefore, many efforts have been made to extend the target base-pair sequence of such DNA ligands. Wilson and coworkers succeeded in selectively targeting a 10 bp sequence with ABP motifs.[6a] Mascareñas and coworkers have recently synthesized a DNA-binding peptide targeting 13 bp, which is composed of the DNA-binding fragments of two transcription factors connected by a hook.[7] Iverson and coworkers have developed threading polyintercalators, composed of naphthalene diimide intercalators and minor groove and major groove binding aliphatic linkers, which target 14 bp and 22 bp.[8] This 22 bp sequence was at that time the longest site recognized by a synthetic, non-nucleic-acid-based DNA binding molecule. Regarding Py–Im polyamides, a linear Py–Im polyamide monomer forming a homodimer with two molecules, and a cysteine-derived Py–Im polyamide dimer were both reported to bind to 16 bp sequences.[21] Furthermore, tandem hairpin motifs composed of several hairpin moieties with connecting hinge segments have also been demonstrated.[22,23]

Recently, our group has designed tandem hairpin Py–Im polyamides specific for the human telomere repetitive sequences 5′–$(TTAGGG)_n$–3′.[23] The human telomere is composed of tandem repeats of 5′–TTAGGG–3′ and the protein

complex called shelterin including TRF1, TRF2, Rap1, TIN2, TPP1 and POT1. The telomere protects the ends of the chromosomes from nucleolytic degradation and DNA recombination and thus is important in chromosome stability.[24] Human telomere DNA has a duplex region and a single-stranded 3′ overhang, which is suggested to form G-quadruplexes or t-loops.[25] In normal mammalian cells, the number of telomere repeats decreases with cell division and is related to the aging process and cancer.[24a,c,f] Therefore, the telomere length is an important biomarker in studies of these processes, and many techniques to visualize telomeres and to measure telomere length have been reported.[15b] Fluorescence in situ hybridization (FISH) using fluorescent PNA probes has been widely used, but this method normally denatures the DNA under the harsh conditions required for efficient probe binding to telomeres and can lead to destruction of telomere structures.[26] In contrast, Py–Im polyamides can bind to target sequences under mild conditions and thus can visualize target regions without destroying their structures.[23e] Based on a previous report of visualization of telomeres and estimation of their length in chemically fixed cells with Texas Red-labeled Py–Im polyamide probes,[23a] our group developed a simple method for synthesis of Py–Im polyamide probes labeled with 5–carboxytetramethylrhodamine (TAMRA) that bound to 12 bp of the human telomere sequences. The probes were named tandem dimer **TDi59-A** and **TAMRA TDi59-A** (Figure 1) and were synthesized from a fluorenylmethyloxycarbonyl (Fmoc) building block comprising the sequence from (R)-Dab to the N-terminus of a hairpin moiety.[23b-d] Costaining with the fluorescent Py–Im polyamide probe and anti-TRF1 antibody under mild conditions confirmed that fluorescently labeled **TDi59-A** could specifically visualize human telomeres in chemically fixed cells and that telomere length is related to the amount of shelterin.[23a,b] To decrease the background signals from nonspecific binding in telomere staining, we also developed **TAMRA TDi59-B**, containing hinge B lengthened by one methylene unit,[23c] and tandem trimer **TAMRA TTri59-A**, whose target sequence was extended to 18 bp, the longest specific sequence reported for Py–Im polyamides.[23d] Furthermore the Py–Im polyamide probes can also visualize telomeres in tissue sections.[23e]

In this study, we have synthesized, using an automatic solid-phase peptide synthesizer, novel tandem tetramer Py–Im polyamides **TTet59-A** and **-B** (Figure 1) targeting 24

Scheme 1. **Solid–phase Synthesis of Tandem Tetramer Py–Im Polyamides and the Structure of Building Block 1.**



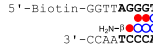
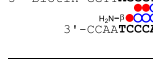
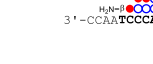
Reagents and conditions: (i) 20% piperidine, DMF; (ii) Fmoc-PyIm-CO_2H, HCTU, DIEA, NMP; (iii) 20% piperidine, DMF; (iv) Fmoc-Py-CO_2H, HCTU, DIEA, NMP; (v) 20% piperidine, DMF; (vi) Fmoc-Py-CO_2H, HCTU, DIEA, NMP; (vii) 20% piperidine, DMF; (viii) **1**, HCTU, DIEA, NMP; (ix) 20% piperidine, DMF; (x) Fmoc-mini-PEG or *N*-Fmoc-Amido-dPEG$_2$ Acid, HCTU, DIEA, NMP; (xi) 20% piperidine, DMF; (xii) Fmoc-PyIm-CO_2H, HCTU, DIEA, NMP; (xiii) 20% piperidine, DMF; (xiv) Fmoc-Py-CO_2H, HCTU, DIEA, NMP; (xv) 20% piperidine, DMF; (xvi) Fmoc-Py-CO_2H, HCTU, DIEA, NMP; (xvii) 20% piperidine, DMF; (xviii) **1**, HCTU, DIEA, NMP; (xix) 20% piperidine, DMF; (xx) 3,3′-diamino-*N*-methyldipropylamine, 45 °C; (xxi) 5–carboxytetramethylrhodamine succinimidyl ester or EZ-Link NHS-PEG$_{12}$-Biotin, DIEA, DMF.

bp in the human telomere 5′–(TTAGGG)$_n$–3′ repeats. These Py–Im polyamides are composed of four hairpin units and three hinges, either A or B. To our knowledge, 24 bp is the new record for the longest binding site of synthetic, non-nucleic-acid-based, sequence-specific DNA-binding molecules. Surface plasmon resonance (SPR) analysis was performed to assess the binding affinity to four telomere repeats of the two tandem tetramer Py–Im polyamides and tandem trimer **TTri59-A**. After conjugation of **TTet59-A** and **-B** with TAMRA to obtain **TAMRA TTet59-A** and **-B**, we stained telomeres in chemically fixed cells to compare the new Py–Im polyamide probes with those previously reported and found that the background signals derived from nonspecific binding were much lower for the tetramers than for previously reported Py–Im polyamide probes. Finally, **TDi59-A** and **TTet59-B** were conjugated to biotin, allowing us to assess their binding to telomeric repeats in the whole genome using high-throughput sequencing technology.[16f,g,27]

Results and Discussion

Synthesis of Py–Im Polyamides. As shown in Scheme 1, tandem tetramer Py–Im polyamides **TTet59-A** and **-B** were synthesized based on the reported methodology, using building block **1** corresponding to Dab and three-ring Py and Im beside the N-terminal.[23b-d] In detail, machine-assisted Fmoc solid-phase peptide synthesis (SPPS) from Fmoc-β-Ala-Wang resin was performed. Fmoc-protected resin was treated with 20% piperidine/*N*,*N*-dimethylformamide (DMF) for deprotection, followed by coupling with the next Fmoc block (Fmoc-Py-OH, Fmoc-PyIm-OH,[28] **1**,[23b-d] Fmoc-mini-PEG or *N*-Fmoc-amido-dPEG2 Acid) activated with *N*,*N*-diisopropylethylamine (DIEA) and *N*,*N*,*N*,*N*-tetramethyl-O-(6-chloro-1H-benzotriazol-1-yl) uronium hexafluorophosphate (HCTU) in *N*-methyl-2-pyrrolidone (NMP). After these two procedures were repeated 19 times, the N-terminal Fmoc-protected amino group was deprotected with 20% piperidine and then cleaved from the resin with 3,3′-diamino-*N*-methyldipropylamine at 45 °C to produce **TTet59-A** and **-B** with 2.7% and 7.7% yield, respectively. The resulting Py–Im polyamides were coupled with 5-carboxytetramethylrhodamine or biotin succinimidyl ester in DIEA and DMF to produce the fluorescent probes **TAMRA TTet59-A** and **-B** and the biotinylated Py–Im polyamides **Biotin TDi59-A** and **Biotin TTet59-B**. Analytical high-performance liquid chromatography (HPLC) profiles and electrospray ionization time-of-flight mass (ESI-TOF-MS) spectrometry spectra are shown in Figures S1 and S2, respectively.

Table 1. **Binding Affinities of Polyamides TTet59–A, TTet59–B and TTri59–A against Match Sequence (ODN-1) Calculated with SPR.**

	k_a ($M^{-1}s^{-1}$)	k_d (s^{-1})	K_D (M)
5′-Biotin-GGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG$_{Tn}$  3′-CCAATCCCAATCCCAATCCCAATCCCAATCC$^{T}_{T}$ TTet59-A	5.7 x 10^4	4.3 x 10^{-4}	7.5 x 10^{-9}
5′-Biotin-GGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG$_{Tn}$  3′-CCAATCCCAATCCCAATCCCAATCCCAATCC$^{T}_{T}$ TTet59-B	4.1 x 10^4	1.6 x 10^{-4}	4.0 x 10^{-9}
5′-Biotin-GGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG$_{Tn}$  3′-CCAATCCCAATCCCAATCCCAATCCCAATCC$^{T}_{T}$ TTri59-A	7.9 x 10^4	4.8 x 10^{-4}	6.1 x 10^{-9}

Comparative Analysis of Binding to Match Sequence. Binding affinities of tetramers **TTet59-A** and **-B** and tandem trimer **TTri59-A** to telomere repeats were assessed with SPR.[23c,d,29] 5′-Biotinylated ODN-1 (5′-biotin-GGTTAGGGTTAGGGTTAGGGTTAGGGTTTTCCTAACCCTAACCCTAACCCTAACCCTAACC-3′) containing four telomeric repeats was immobilized to a sensor chip through a biotin–streptavidin interaction, and then the trimer and tetramers were passed over the DNA on the sensor chip, and sensorgrams (shown in Figure S3) and values (summarized in Table 1) were obtained. The underlined bases are the binding site of these three types of Py–Im polyamides. Tetramer **TTet59-A** showed strong binding affinity to ODN-1 (K_D = 7.5 nM), comparable to other types of Py–Im polyamides.[9] Trimer **TTri59-A** showed a slightly higher association constant and stronger binding affinity than **TTet59-A** (K_D = 6.1 nM). As suggested previously,[23d] a major reason for this result is that the tetramer's higher steric hindrance influenced its accessibility to the match site. Of the three Py–Im polyamides, **TTet59-B** showed the strongest binding affinity (K_D = 4.0 nM), attributed to a much slower dissociation, suggesting that the hinge region prevented the polyamide's dissociation.

Visualization of Human Telomeres to Compare Py–Im Polyamide Probes. To compare the abilities of Py–Im polyamide probes to stain telomeres specifically, we doubly stained human HeLa 1.3 cell spreads with 4′,6-diamidino-2-phenylindole (DAPI) and five Py–Im polyamide probes: **TAMRA TDi59-A** and **-B**, **TAMRA TTri59-A**, and **TAMRA TTet59-A** and **-B**.[23] Cell spreads are used for clinical karyotyping tests and are prepared from colcemid-treated mitotic cells fixed with MeOH/AcOH. The resulting images of chromosomes stained with 75 nM Py–Im polyamide probes are shown in Figures 2 and S4A. Chromosomal regions and nuclei were visualized with DAPI. Each Py–Im polyamide probe showed two foci at the ends of every chromosome, suggesting that all probes, including the new tandem tetramer probes **TAMRA TTet59-A** and **-B**, could stain telomeres specifically. As the number of targeted base pairs increased, the fluorescence of the polyamide probes became lower, presumably because increasing the size of the polyamide probes decreased the number of molecules binding to telomeres, but the background signals derived from nonspecific binding of polyamides decreased. When treating cells with **TAMRA TDi59-A** or **-B** or **TAMRA TTri59-A**, we could easily observe the shape of chromosomes from the background signals of polyamides, and stronger background signals for **TAMRA TDi59-A** in particular were detected from the whole chromosomes. However, for treatment with **TAMRA TTet59-A** or **-B**, the intensity of background signals was less, showing that extending the target base pair sequence significantly improves telomere specificity in cells. In addition, **TAMRA TTet59-B** had less background signal along the chromosomes than **TAMRA TTet59-A**, suggesting that the change

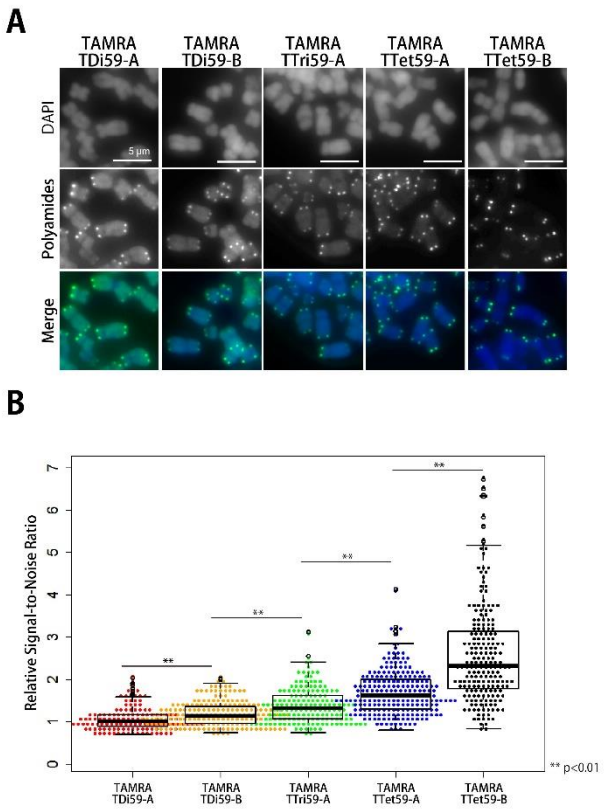


Figure 2. Telomere staining of HeLa 1.3 cell spreads with 75 nM fluorescent Py–Im polyamide probes. (A) The images of HeLa 1.3 cell spreads stained with DAPI (first row) and the fluorescent polyamide probes (second row). The merged images are shown in the third row. Original images are shown in Figure S4A. (B) Distribution of relative signal-to-noise (S/N) ratio of telomeric foci in the images.

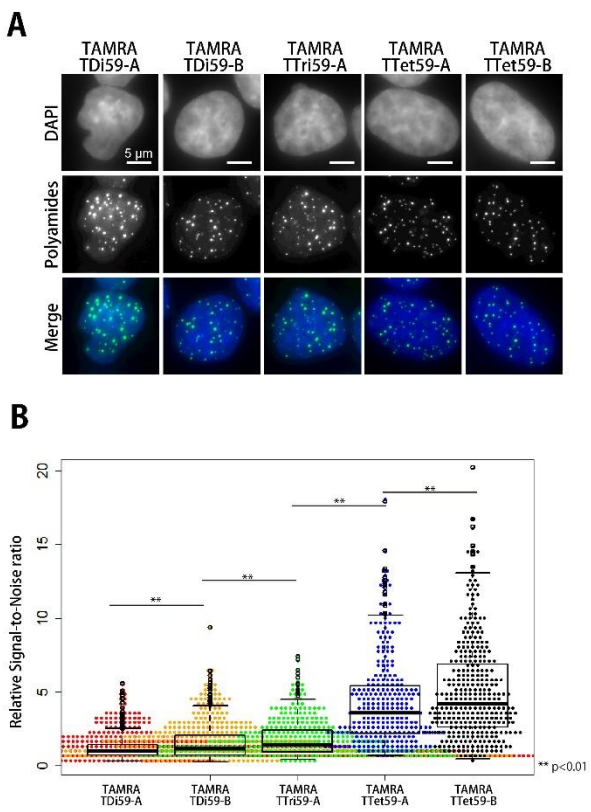


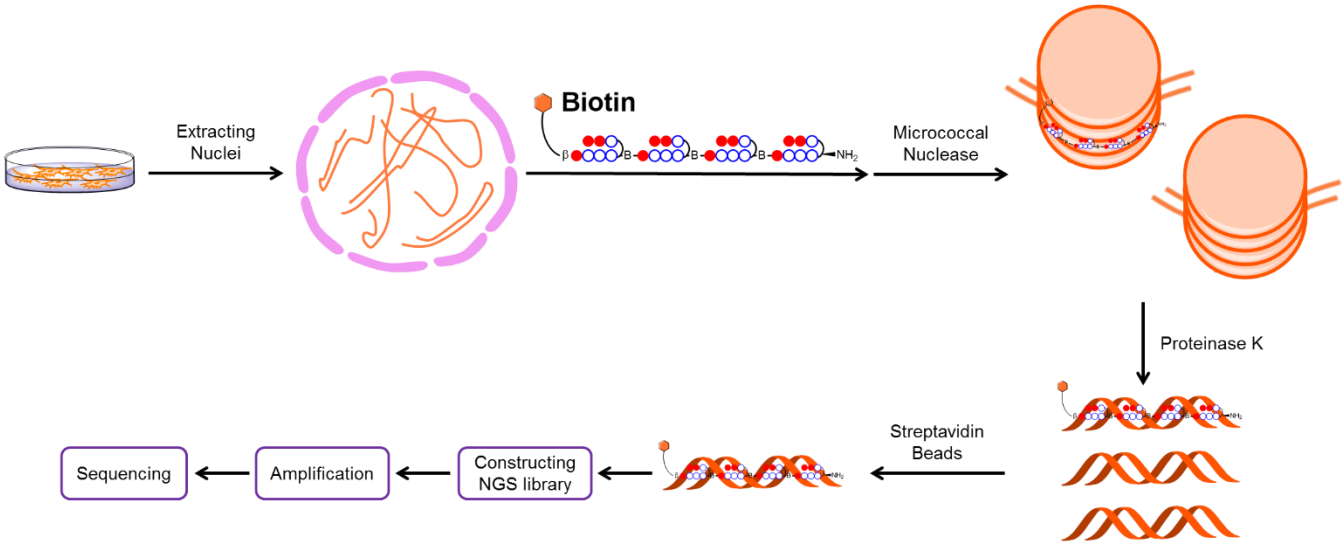
Figure 3. Telomere staining of HeLa 1.3 cells with 15 nM fluorescent Py–Im polyamide probes. (A) The images of HeLa 1.3 cells stained with DAPI (first row) and fluorescent polyamide probes (second row). The merged images are shown in the third row. Original images are shown in Figure S4B. (B) Distribution of signal-to-noise (S/N) ratio of telomeric foci in the images.

of hinge from A to B is also effective in improving specificity; this was also shown by the comparison between **TAMRA TDi59-A** and **-B**.[23c] As shown in Figure 2B, **TAMRA TTet59-B** had the highest relative signal-to-noise (S/N) ratio. These results demonstrated that increasing the number of target base pairs to 24 and changing the hinge from A to B suppressed the nonspecific binding and allowed us to observe telomeres in HeLa 1.3 cell spreads with high specificity.

Images of telomere staining of HeLa 1.3 cells fixed in formaldehyde (Figures 3 and S4B), a commonly used fixation method in cell biology, and the quantitative analysis of S/N ratios (Figure 3B) suggested that the results were consistent with those of cell spreads staining. Many sharp foci from the telomeric regions were observed in the nuclei of cells.[23a,b] As reported previously, **TAMRA TDi59-B** and **TAMRA TTri59-A** showed lower background signals than **TAMRA TDi59-A**.[23c,d] The background signals of the two tandem tetramer probes **TAMRA TTet59-A** and **-B** were lower and their S/N ratios were higher than the other three probes. As can be seen from the polyamide signals of **TAMRA TTet59-A** and **-B** shown in Figure 3, the shapes of the whole nuclei were less distinct. In the images of formaldehyde-fixed cells, **TAMRA TTet59-B** also showed the highest S/N ratio. These results also demonstrated that **TAMRA TTet59-A** and **-B** could visualize telomeres with higher specificity in the fixed cells or formaldehyde-fixed cells and that **TAMRA TTet59-B** targeting 24 bp and composed of four hairpins and three B hinges had the highest telomere specificity of the fluorescent Py–Im polyamide-based telomere probes.

High-throughput Sequencing of the Binding Sites of Py–Im Polyamide Probes for Telomeric Repeats. Images of chromosome spreads treated with the Py–Im polyamide probes (Figure 2) suggested that they specifically stained telomeres. To assess their efficient binding to telomeric repeats in a biologically active, histone-packed, chromatinized genome, we performed high-throughput sequencing of DNA bound to **Biotin TDi59-A** and **Biotin TTet59-B,** based on the reported procedure.[16f,g] The workflow is shown in Scheme 2. In brief, nuclei were extracted from BJ fibroblast cells and then incubated with 400 nM **Biotin TDi59-A** or **Biotin TTet59-B** for 16 h at 4 °C. After this, the extracted nuclei were treated with micrococcal nuclease to produce mononucleosomes to generate the optimally sized DNA for construction of a sequencing library. Nuclear proteins were digested with proteinase K, and then DNA bound to the biotinylated Py–Im polyamide

Scheme 2. **Extraction and High-throughput Sequencing of Genomic DNA Bound by Biotinylated Py–Im Polyamide Probes.**



probes was extracted through a biotin–streptavidin interaction on streptavidin-coated magnetic beads (C1 beads). The extracted and purified DNA was ligated with sequencing platform-specific adapters and amplified with a polymerase chain reaction to obtain sufficient amounts for a sequencing library. The qualified libraries were sequenced using Ion Proton Sequencing system. Sequenced and filtered reads were aligned to human reference genome hg38. MACS peaks were called to identify the differentially enriched regions compared with control data.[30a]

Figure S5 shows a representative result indicating the enriched regions in chromosomes 1 and 18, and a close-up of the enriched telomeric repeats from a non-treated control sample and after **Biotin TDi59-A** and **Biotin TTet59-B** treatment is shown in Figure 4. In both left and right arm termini of chromosomes 1 and 18, we observed significant enrichments with both polyamides compared with the control. The enrichment was consistent with telomeric repeat sequences (TTAGGG/CCCTAA) of the reference genome, suggesting that we could successfully validate the binding of **Biotin TDi59-A** and **Biotin TTet59-B** to telomeric repeats with high-throughput sequencing. Other mild enrichments appeared to be the results of partial or non–specific binding.

The enrichments in the termini of each chromosome are shown in Figure S6. Duplicate experiments ensured the consistency of enrichments in chromosomal termini. However, we could not quantify the enrichment level on Chr 3R, 5R, 6L, 6R, 8L, 8R, 9R, 11L, 13L, 14L, 14R, 15L, 16R, 17L, 19L, 20L, 20R, 21L and 22L (L- left arm, R- right arm) chromosomal termini. This may be because of lacking telomeric repeat sequences information on the above mentioned chromosomes in the human reference genome hg38 data. Even though chromosomes 13, 14, 15, 21 and 22 lack the telomeric repeats information at the left termini presumably because of rDNA whose sequences are not included in the human reference genome hg38, we could observe enrichments from ~16000000 bp, ~16000000 bp, ~17000000 bp, ~5010000 bp and ~10510000 bp, respectively. These results show that **Biotin TDi59-A** and **Biotin TTet59-B** could effectively recognize and bind to telomeric repeats.

To identify high affinity binding motif of **Biotin TDi59-A** and **Biotin TTet59-B** from the enrichments, we used the Homer motif analysis program with the enriched peaks.[30b] One of the highly enriched motifs corresponded to telomeric 5′–$(TTAGGG)_n$–3′ repeats with a *p* value of e^{-8595} and e^{-100} for **Biotin TDi59-A** and **Biotin TTet59-B**, respectively. (Figure 4B). This result also suggests efficient binding of **Biotin TDi59-A** and **Biotin TTet59-B** to telomeres.

Finally, to compare the recognition of telomeric repeats by **Biotin TDi59-A** with that by **Biotin TTet59-B**, the ratio of reads containing ≥ 4 TTAGGG/CCCTAA repeats to whole sequenced reads was calculated. As shown in Figure 4C, the ratio for **Biotin TDi59-A** was 18.2 times higher than the nontreated control, showing that **Biotin TDi59-A** could enrich telomeric DNA. The ratio for **Biotin TTet59-B** was 1.9 times higher than that for **Biotin TDi59-A**, suggesting that **TTet59-B** exhibited better recognition ability toward telomeric repeats. This result is also consistent with the higher telomere-specific staining of **TAMRA TTet59-B**.

High-throughput sequencing confirmed the binding of **Biotin TDi59-A** and **Biotin TTet59-B** to telomeric repeats in the whole genome. In contrast, foci of fluorescently labeled Py–Im polyamide probes were specifically detected at the end of the chromosomes (Figure 2). The principal reason for this is that accumulation of the Py–Im polyamides on long telomeric repeats in telomere regions was higher than that in nonspecific regions, and therefore much stronger fluorescence was detected at the termini of chromosomes than from other chromosomal regions.

Conclusion

In this study, we have synthesized novel tandem tetramer Py–Im polyamides composed of 4 hairpins and 3 hinges that targeted 24 bp of the human telomere sequences. This is a new record for the longest binding site of synthetic, non-nucleic-acid-based, sequence-specific

DNA-binding molecules. SPR analysis revealed that tandem tetramers could bind to four telomeric repeats with nanomolar K_D values. TAMRA-labeled tandem tetramer Py–Im polyamide probes could stain clearly the telomere foci in chemically fixed cells. Compared with previously reported tandem dimer and trimer Py–Im polyamides, the tandem tetramers had higher specificity for telomeres in cells, and **TAMRA TTet59-B** containing hinge B had the highest specificity. Furthermore, high-throughput sequencing of chromatin pulled down by biotin-labeled Py–Im polyamide probes confirmed their high recognition and binding toward telomeric DNA. To the best of our knowledge, this is the first report of high-throughput sequencing of DNA pulled down by a synthetic DNA binder targeting human telomeres. To allow the application of these tetramers, our group continues to optimize further their chemical structures and to develop new functions.

ASSOCIATED CONTENT

Supporting Information. Materials and Methods, Mass spectra and HPLC profiles of tandem tetramers, SPR sensorgrams, Cell images, Some enrichment data obtained from high throughput sequencing. This information is available free of charge via the internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

(H. S.) hs@kuchem.kyoto-u.ac.jp
(K. M.) kmaeshim@nig.ac.jp
(T. B.) bando@kuchem.kyoto-u.ac.jp

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

We thank Dr T. de Lange (Rockefeller University) for the gift of HeLa 1.3 cells, Dr S. Sugano (Kyoto University) for helpful discussion about high throughput sequencing, S. Asamitsu (Kyoto University) for assistance in SPR data analysis, S. Sato (Kyoto University) for culturing BJ cells. Y. K. is thankful to JSPS research fellowship for young scientists (DC1). This work was supported by Basic Science and Platform Technology Program for Innovative Biological Medicine to H. S., JSPS–NSF International Collaborations in Chemistry (ICC) to H. S., NIG

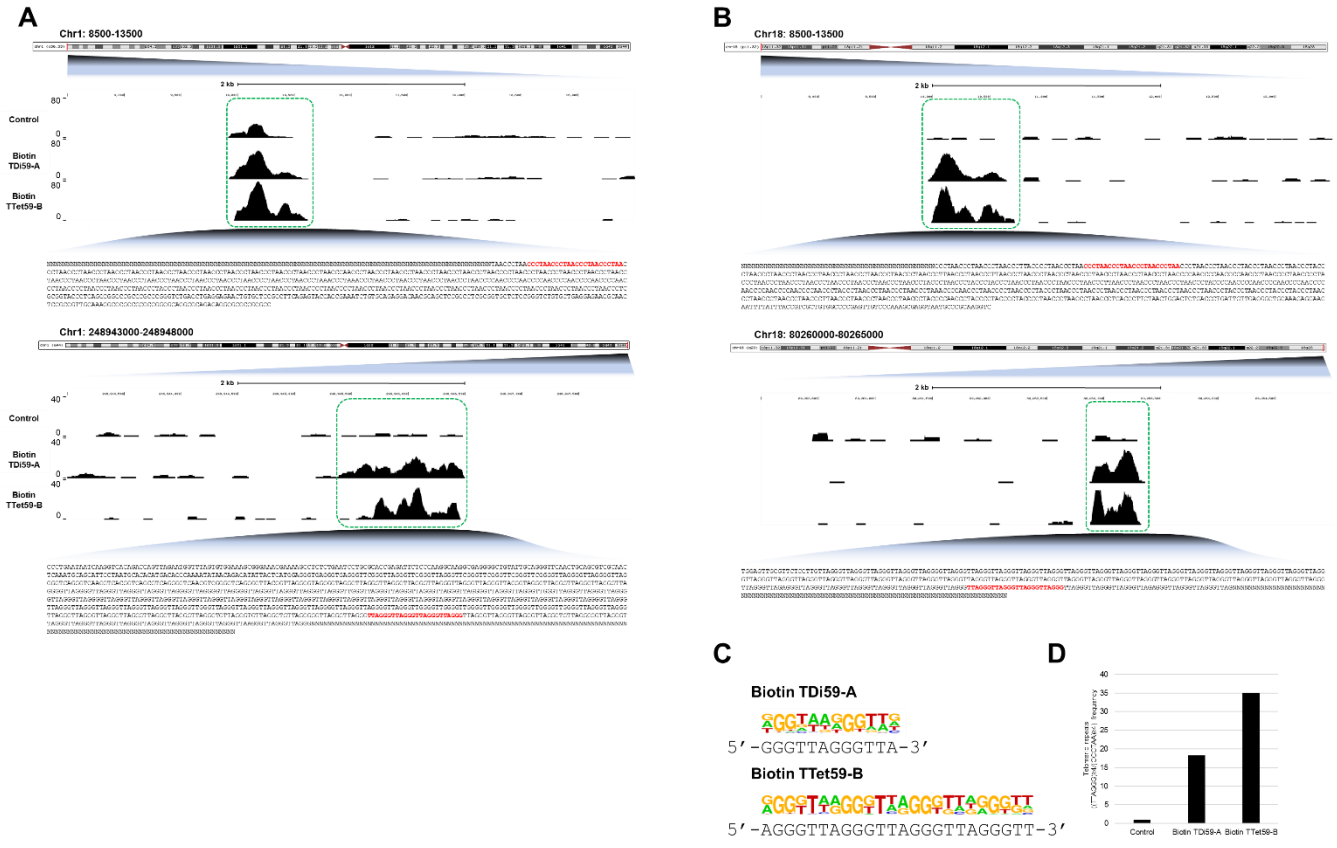


Figure 4. Assessment of efficient binding of **Biotin TDi59–A** and **Biotin TTet59–B** to telomeric repeats with high–throughput sequencing. Mapping of the biotinylated Py–Im polyamide probes binding and enriched sites in both termini of chromosomes 1 and 18 are shown in (A) and (B) respectively. (C) Binding motifs of the probes corresponding to telomeric repeats identified in human genomic enriched sequences. (D) The ratio of sequenced reads containing ≥ 4 telomeric repeats to whole reads as normalized to that from the control.

Collaborative Research Program (2015–B6) to K. M. and H. S., JST–CREST to K. M. and JSPS KAKENHI Grant number 24225005 to H. S., 24310155 to T. B. and 15J00928 to Y. K.

REFERENCES

(1) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. *Science* **2012**, *337*, 816–821.

(2) Klug, A. *Annu. Rev. Biochem.* **2010**, *79*, 213–231.

(3) Boch, J.; Scholze, H.; Schornack, S.; Landgraf, A.; Hahn, S.; Kay, S.; Lahaye, T.; Nickstadt, A.; Bonas, U. *Science* **2009**, *326*, 1509–1512.

(4) (a) Duca, M.; Vekhoff, P.; Oussedik, K.; Halby, L.; Arimondo, P. B. *Nucleic Acids Res.* **2008**, *36*, 5123−5138. (b) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645–650.

(5) (a) Nielsen, P. E. *Chem. Biodivers.* **2010**, *7*, 786–804. (b) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991**, *254*, 1497 –1500.

(6) (a) Liu, Y.; Chai, Y.; Kumar, A.; Tidwell, R. R.; Boykin, D. W.; Wilson, W. D. *J. Am. Chem. Soc.* **2012**, *134*, 5290−5299. (b) Paul, A.; Nanjunda, R.; Kumar, A.; Laughlin, S.; Nhili, R.; Depauw, S.; Deuser, S. S.; Chai, Y.; Chaudhary, A. S.; David–Cordonnier, M. H.; Boykin, D. W.; Wilson, W. D. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 4927−4932.

(7) Rodríguez, J.; Mosquera, J.; García-Fandiño, R.; Vázquez, M. E.; Mascareñas, J. L. *Chem. Sci.* **2016**, *7*, 3298–3303.

(8) (a) Holman, G. G.; Zewail–Foote, M.; Smith, A. R.; Johnson, K. A.; Iverson, B. L. *Nat. Chem.* **2011**, 3, 875−881. (b) Smith, A. R.; Iverson, B. L. *J. Am. Chem. Soc.* **2013**, *135*, 12783–12789.

(9) (a) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215−2235. (b) Dervan, P. B.; Edelson, B. S. *Curr. Opin. Struc. Biol.* **2003**, *13*, 284−299. (c) Dervan, P. B.; Doss, R. M.; Marques, M. A. *Curr. Med. Chem.: Anti–Cancer Agents.* **2005**, 5, 373−387. (d) Bando, T.; Sugiyama, H. *Acc. Chem. Res.* **2006**, *39*, 935−944. (e) Blackledge, M. S.; Melander, C. *Bioorg. Med. Chem.* **2013**, *21*, 6101−6114. (f) Trauger, J. W.; Baird, E. E.; Dervan P. B. *Nature* **1996**, *382*, 559−561. (g) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468−471.

(10) (a) Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382−1391. (b) Swalley, S. E.; Eldon E. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1113−1120.

(11) (a) Mrksich, M.; Parks, M. E.; Dervan. P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983−7988. (b) de Clairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909−7916.

(12) (a) Herman, D. M.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1121−1129. (b) Chenoweth, D. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 13175−13179. (c) Chenoweth, D. M.; Dervan, P. B. *J. Am. Chem. Soc.* **2010**, *132*, 14521−14529. (d) Morinaga, H.; Bando, T.; Takagaki, T.; Yamamoto, H.; Hashiya, K.; Sugiyama, H. *J. Am. Chem. Soc.* **2011**, *133*, 18924−18930. (e) Li, B. C.; Montgomery, D. C.; Puckett, J. W.; Dervan, P. B. *J. Org. Chem.* **2013**, *78*, 124−133.

(13) (a) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147−6152. (b) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219−6226.

(14) (a) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141−6146. (b) Wurtz, N. R.; Turner. J. M.; Baird, E. E.; Dervan, P. B. *Org. Lett.* **2001**, *3*, 1201−1203. (c) Fan, L.; Yao, G.; Pan, Z.; Wu, C.; Wang, H. S.; Burley, G. A.; Su, W. *Org. Lett.* **2015**, *17*, 158−161.

(15) (a) Vaijayanthi, T.; Bando, T.; Pandian, G. N.; Sugiyama, H. *ChemBioChem* **2012**, *13*, 2170−2185. (b) Boutorine, A. S.; Novopashina, D. S.; Krasheninina, O. A.; Nozeret, K.; Venyaminova, A. G. *Molecules* **2013**, *18*, 15357−15397. (c) Nozeret, K.; Loll, F.; Escudé, C.; Boutorine, A. S. *ChemBioChem* **2015**, *16*, 549−554.

(16) (a) Anandhakumar, C.; Kizaki, S.; Bando, T.; Pandian, G. N.; Sugiyama, H. *ChemBioChem* **2015**, *16*, 20–38. (b) Meier, J. L.; Yu, A. S.; Korf, I.; Segal, D. J.; Dervan, P. B. *J. Am. Chem. Soc.* **2012**, *134*, 17814−17822. (c) Kang, J. S.; Meier, J. L.; Dervan, P. B. *J. Am. Chem. Soc.* **2014**, *136*, 3687−3694. (d) Erwin, G. S.; Bhimsaria, D.; Eguchi, A.; Ansari, A. Z. *Angew. Chem. Int. Ed.* **2014**, 53, 10124−10128. (e) Anandhakumar, C.; Li, Y.; Kizaki, S.; Pandian, G. N.; Hashiya, K.; Bando, T.; Sugiyama, H. *ChemBioChem.* **2014**, *15*, 2647−2651. (f) Chandran, A.; Syed, J.; Taylor, R. D.; Kashiwazaki, G.; Sato, S.; Hashiya, K.; Bando, T.; Sugiyama, H. *Nucleic Acids Res.* **2016**, *44*, 4014−4024. (g) Chandran, A.; Syed, J.; Li, Y.; Sato, S.; Bando, T.; Sugiyama, H. *ChemBioChem.* **2016**, DOI: 10.1002/cbic.201600274.

(17) (a) Murty, M. S. R. C.; Sugiyama, H. *Biol. Pharm. Bull.* **2004**, 27, 468−474. (b) Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202−205. (c) Lai, Y.–M.; Fukuda, N.; Ueno, T.; Kishioka, H.; Matsuda, H.; Saito, S.; Matsumoto, K.; Ayame, H.; Bando, T.; Sugiyama, H.; Mugishima, H.; Serie, K. *J. Pharmacol. Exp. Ther.* **2005**, *315*, 571–575. (d) Hiraoka, K.; Inoue, T.; Taylor, R. D.; Watanabe, T.; Koshikawa, N.; Yoda, H.; Shinohara, K.; Takatori, A.; Sugimoto, K.; Maru, Y.; Denda, T.; Fujiwara, K.; Balmain, A.; Ozaki, T.; Bando, T.; Sugiyama, H.; Nagase, H. *Nat. Commun.* **2015**, *6*, 6706.

(18) (a) Xiao, X.; Yu, P.; Lim, H. S.; Sikder, D.; Kodadek, T. *Angew. Chem. Int. Ed.* **2007**, *46*, 2865−2868. (b) Patel, S.; Jung, D.; Yin, P.; Carlton, P.; Yamamoto, M.; Bando, T.; Sugiyama, H.; Lee, K. B. *ACS Nano.* **2014**, *8*, 8959−8967.

(19) (a) Han, L.; Pandian, G. N.; Junetha, S.; Sato, S.; Chandran, A.; Taniguchi, J.; Saha, A.; Bando, T.; Nagase, H.; Sugiyama, H. *Angew. Chem. Int. Ed.* **2013**, 52, 13410−13413. (b) Pandian, G. N.; Taniguchi, J.; Junetha, S.; Sato, S.; Han, L.; Saha, A.; Anandhkumar, C.; Bando, T.; Nagase, H.; Vaijayanthi, T.; Taylor, R. D.; Sugiyama, H. *Sci. Rep.* **2014**, *4*, 3843. (c) Han, L.; Pandian, G, N.; Chandran, A.; Sato, S.; Taniguchi, J.; Kashiwazaki, G.; Sawatani, Y.; Hashiya, K.; Bando, T.; Xu, Y.; Qian, X.; Sugiyama, H. *Angew. Chem. Int. Ed.* **2015**, *54*, 8700–8703.

(20) Singh, I.; Wendeln, C.; Clark, A. W.; Cooper, J. M.; Ravoo, B. J.; Burley, G. A. *J. Am. Chem. Soc.* **2013**, *135*, 3449−3457.

(21) (a) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534−3535. (b) Yamamoto, M.; Bando, T.; Morinaga, N.; Kawamoto, Y.; Hashiya, K.; Sugiyama, H. *Chem. Eur. J.* **2014**, *20*, 752−759.

(22) (a) Herman, D. M.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1999**, *5*, 975−983. (b) Kers, I.; Dervan, P. B. *Bioorg. Med. Chem.* **2002**, *10*, 3339−3349. (c) Schaal, T. D.; Mallet, W. G.; McMinn, D. L.; Nguyen, N. V.; Sopko, M. M.; John, S.; Parekh, B. S. *Nucleic Acids Res.* **2003**, *31*, 1282−1291. (d) Sasaki, S.; Bando, T.; Minoshima, M.; Shinohara, K.; Sugiyama, H. *Chem. Eur. J.* **2008**, *14*, 864−870. (e) Yamamoto, M.; Bando, T.; Kawamoto, Y.; Taylor, R.; Hashiya, K.; Sugiyama, H. *Bioconjugate Chem.* **2014**, *25*, 552−559. (f) Taylor, R, D.; Kawamoto, Y.; Hashiya, K.; Bando, T.; Sugiyama, H. *Chem. Asian J.* **2014**, *9*, 2527−2533.

(23) (a) Maeshima, K.; Janssen, S.; Laemmli, U. K. *EMBO J.* **2001**, *20*, 3218−3228. (b) Kawamoto, Y.; Bando, T.; Kamada, F.; Li, Y.; Hashiya, K.; Maeshima, K.; Sugiyama, H. *J. Am. Chem. Soc.* **2013**, *135*, 16468−16477. (c) Hirata, A.; Nokihara, K.; Kawamoto, Y.; Bando, T.; Sasaki, A.; Ide, S.; Maeshima, K.; Kasama, T.; Sugiyama, H. *J. Am. Chem. Soc.* **2014**, *136*, 11546−11554. (d) Kawamoto, Y.; Sasaki, A.; Hashiya, K.; Ide, S; Bando, T.; Maeshima, K; Sugiyama. H. *Chem. Sci.* **2015**, *6*, 2307−2312. (e) Sasaki, A.; Ide, S.;

Kawamoto, Y.; Bando, T.; Murata, Y.; Shimura, M.; Yamada, K.; Hirata, A.; Nokihara, K.; Hirata, T.; Sugiyama, H.; Maeshima, K. *Sci. Rep.* **2016**, *6*, 29261.

(24) (a) Smogorzewska, A.; de Lange, T. *Annu. Rev. Biochem.* **2004**, *73*, 177–208. (b) Palm, W.; de Lange, T. *Annu. Rev. Genet.* **2008**, *42*, 301–334. (c) Blackburn, E. H. *Angew. Chem. Int. Ed.* **2010**, *49*, 7405–7421. (d) Xu, Y. *Chem. Soc. Rev.* **2011**, *40*, 2719–2740. (e) Nandakumar, J.; Cech, T. R. *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 69–82. (f) Zakian, V. A. *Exp. Cell Res.* **2012**, *318*, 1456–1460. (g) Smogorzewska, A.; van Steensel, B.; Bianchi, A.; Oelmann, S.; Schaefer, M. R.; Schnapp, G.; de Lange, T. *Mol. Cell Biol.* **2000**, *20*, 1659–1668.

(25) (a) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. *Nat. Chem.* **2013**, *5*, 182–186. (b) Doksani, Y.; Wu, J, Y.; de Lange, T.; Zhuang, X. *Cell* **2013**, *155*, 345–356.

(26) Lansdorp, P. M.; Verwoerd, N. P.; van de Rijke, F. M.; Dragowska, V.; Little, M. T.; Dirks, R. W.; Raap, A. K.; Tanke, H. J. *Hum. Mol. Genet.* **1996**, *5*, 685–691.

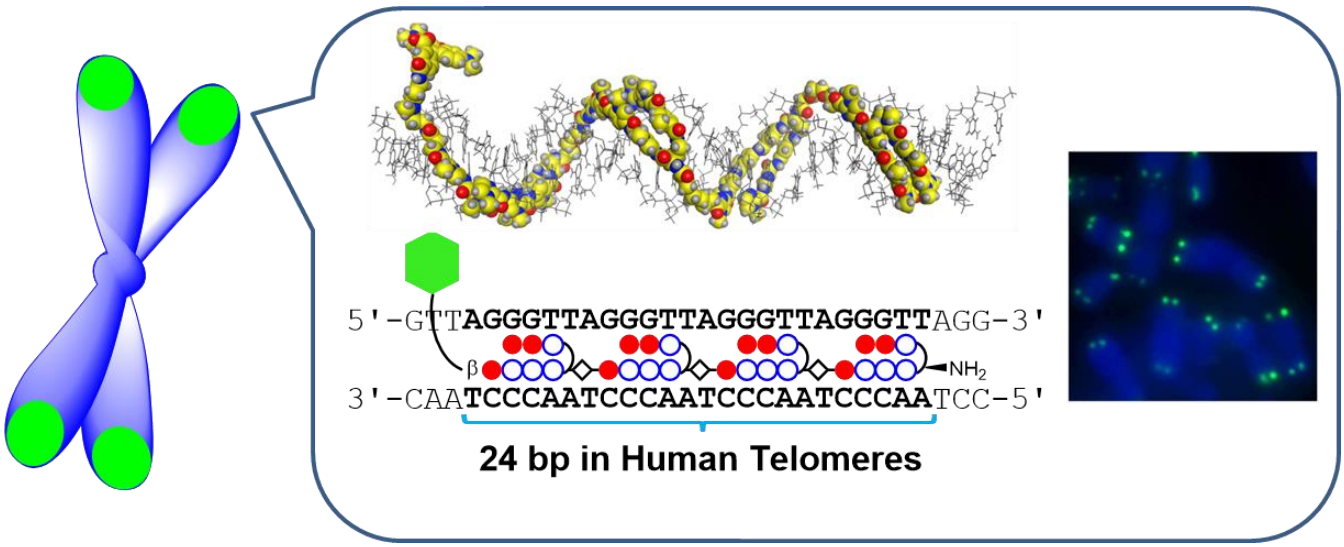
(27) Anders, L.; Guenther, M. G.; Qi, J.; Fan, Z. P.; Marineau, J. J.; Rahl, P. B.; Lovén, J.; Sigova, A. A.; Smith, W. B.; Lee, T. I.; Bradner, J. E.; Young, R. A. *Nat. Biotechnol.* **2014**, *32*, 92–96.

(28) (a) Minoshima, M.; Bando, T.; Sasaki, S.; Fujimoto, J.; Sugiyama, H. *Nucleic Acids Res.* **2008**, *36*, 2889–2894. (b) Wetzler, M.; Wemmer, D. E. *Org. Lett.* **2010**, *12*, 3488–3490.

(29) (a) Lacy, E. R.; Le, N. M.; Price, C. A.; Lee, M.; Wilson, W. D. *J. Am. Chem. Soc.* **2002**, *124*, 2153–2163. (b) Wang, S.; Aston, K.; Koeller, K. J.; Harris, G. D.; Rath, N. P.; Bashkin, J. K.; Wilson, W. D. *Org. Biomol. Chem.* **2014**, *12*, 7523–7536.

(30) (a) Feng, J.; Liu, T.; Qin, B.; Zhang, Y.; Liu, X. S. *Nat. Protoc.* **2012**, *7*, 1728–1740. (b) Heinz, S.; Benner, C.; Spann, N.; Bertolino, E.; Lin, Y. C.; Laslo, P.; Cheng, J. X.; Murre, C.; Singh, H.; Glass, C. K. *Mol. Cell* **2010**, *38*, 576–589.

Insert Table of Contents artwork here



Supporting Information

Targeting 24 bp in Telomere Repeat Sequences with Tandem Tetramer Pyrrole–Imidazole Polyamide Probes

Yusuke Kawamoto[†], Asuka Sasaki[‡], Anandhakumar Chandran[†], Kaori Hashiya[†], Satoru Ide[‡], Toshikazu Bando[*,†], Kazuhiro Maeshima[*,‡], and Hiroshi Sugiyama[*,†,§]

[†]Department of Chemistry, Graduate School of Science, Kyoto University, Sakyo, Kyoto 606–8502, Japan
[‡]Biological Macromolecules Laboratory, Structural Biology Center, National Institute of Genetics, and Department of Genetics, School of Life Science, Graduate University for Advanced Studies (Sokendai), Mishima, Shizuoka 411–8540, Japan.
[§]Institute for Integrated Cell–Material Science (WPI–iCeMS), Kyoto University, Sakyo, Kyoto 606–8501, Japan.

Corresponding authors
(H. S.) hs@kuchem.kyoto-u.ac.jp
(K. M.) kmaeshim@nig.ac.jp
(T. B.) bando@kuchem.kyoto-u.ac.jp

Table of Contents

Materials and Methods

Materials. HPLC analysis for chasing each solution–phase reaction and checking each product was performed on a PU–2089 plus series system (JASCO) using CHEMCOBOND 4.6 mm x 150 mm 5–ODS–H Column (Chemco Plus Scientific Co., Ltd.) in 0.1% TFA in water with acetonitrile as the eluent at a flow rate of 1.0 mL/min and a linear gradient elution of 0 to 50% acetonitrile in 20 min with detection at 254 nm. Purification of Py–Im polyamides were done with reversed–phase HPLC performed on a UV2075 HPLC UV/VIS detector and a PU–2080 plus series system (JASCO) using CHEMCOBOND 4.6 mm x 150 mm 5–ODS–H Column in 0.1% TFA in water with acetonitrile as the eluent at a flow rate of 1.0 mL/min. Collected fractions were lyophilized and then analyzed by ESI–TOF–MS (Bruker). For the collect mass measurement, external mass calibration (m/z: 622.0290 and 922.0098), purchased from Agilent Technologies, was used. Fmoc-mini-PEG and *N*-Fmoc-Amido-dPEG$_2$ Acid were purchased from Peptide International. Fmoc-Py-CO_2H, DMF used for solid–phase synthesis, NMP, trifluoroacetic acid (TFA), and piperidine were from Wako. HCTU was from Peptide Institute, Inc. Fmoc-β-Ala-Wang resin (0.60 mmol/g, 100‑200 mesh) was from Novabiochem. DIEA was from Nacalai Tesque, Inc. DMF used for solution–phase synthesis and for telomere–staining and acetonitrile were from Kanto Chemical Co., Inc. 3, 3'-diamino-*N*-methyldipropyl-amine was from Tokyo Chemical Industry Co., Ltd. 5–carboxytetramethylrhodamine succinimidyl ester and EZ–Link NHS–PEG$_{12}$–Biotin were from Thermo Fisher Scientific Inc. The Fmoc units **1**[1a,b] and Fmoc-PyIm-CO_2H[2] were synthesized using the previous method. **TAMRA TDi59–B** was purchased from HiPep Laboratories. **TAMRA TDi59–A**, **TTri59–A** and **TAMRA TTri59–A** were synthesized following the previous methodology.[1b,c] Solid–phase peptide synthesis was performed on a PSSM–8 (Shimadzu). Concentrations of Py–Im Polyamides were calculated with a Nanodrop ND–1000 spectrophotometer (Thermo Fisher Scientific Inc.) using an extinction coefficient of 9900 $M^{-1}cm^{-1}$ per one pyrrole or imidazole moiety at λ_{max} near 310 nm. The SPR assays were performed using a BIACORE X instrument (GE Healthcare) with 5'–biotinylated DNA oligonucleotides (JBioS), streptavidin–coated sensor chip SA (GE Healthcare) and HBS–EP buffer (GE Healthcare). Cell images were recorded with DeltaVision (Applied Precision). HeLa 1.3 cells were generous gifts of Dr T. de Lange (Rockefeller University) and BJ fibroblast cells were purchased from ATCC. DMEM medium and streptavidin–coated magnetic beads were purchased from Invitrogen. Normal goat serum (NGS) was from Millipore. Paraphenylene diamine, Triton X–100 and protease inhibitor cocktail were from Sigma.

DAPI was from Roche. MNase was from TaKaRa Bio Inc. Nonidet P–40 was from Nacalai Tesque, Inc. PCR purification kit was from Qiagen. Quality and quantity check of sequencing library was performed with Agilent 2100 Bioanalyzer and High sensitivity BioAnalyzer kit (Agilent Technologies). High throughput sequencing was performed with Ion Proton™ Sequencer (Thermo Fisher Scientific Inc.). All other solvents and materials were from standard suppliers (highest quality available).

General Procedures of Solid–phase Synthesis of Tandem Tetramer Py–Im Polyamides. Following the reported condition,[1c] each Fmoc solid–phase peptide synthesis (SPPS) of tandem tetramers was performed with a computer–assisted operation system under low humidity (< 20%). Fmoc units in each step were as follows: Fmoc-Py-CO_2H (77 mg, 0.21 mmol), Fmoc-PyIm-CO_2H (70 mg, 0.14 mmol), Fmoc-D-Dab(ImImPy)-OH (**1**) (70 mg, 0.10 mmol), Fmoc-mini-PEG (81 mg, 0.21 mmol) and *N*-Fmoc-Amido-dPEG$_2$ Acid (84 mg, 0.21 mmol). Before SPPS, each Fmoc unit was dissolved in NMP (1.0 mL) with HCTU and the amount of HCTU for Fmoc-Py-CO_2H, Fmoc-PyIm-CO_2H, **1**, Fmoc-mini-PEG and *N*-Fmoc-Amido-dPEG$_2$ Acid was 88 mg (0.21 mmol), 60 mg (0.15 mmol), 40 mg (0.097 mmol), 88 mg (0.21 mmol) and 88 mg (0.21 mmol), respectively. Fmoc-β-Ala-Wang resin swelled with NMP was loaded in the reaction vessel.

Procedures were as follows: twice deblocking for 4 min with 20 % piperidine/ DMF (500 μL), addition of 10% DIEA/NMP (364 μL, DIEA was 0.21 mmol) to Fmoc unit-HCTU mixture for activation, coupling for 60 min. Five times washing with DMF were performed during each step. After the last coupling, the deblocking of amino group at γ-turn was performed. All couplings were carried out with a single-coupling cycle. All lines were purged with solution transfers and bubbled by nitrogen gas for stirring the resin.

After SPPS, Py–Im polyamides on resin were cleaved with 3,3’-diamino-*N*-methyldipropylamine at 45 °C for 3 h. The resin was removed by filtration and washed thoroughly with dichloromethane, and the filtrate was concentrated *in vacuo*. The residue was dissolved in 1.0–2.0 mL dichloromethane-methanol mixture and then more than 10-fold volume of diethyl ether was added, followed by the centrifuge and removal of the supernatant. This process was repeated a few times until white precipitation was obtained. The obtained crude products were purified by reversed–phase HPLC in 0.1% TFA in water with acetonitrile as the eluent at a flow rate of 1.0 mL/min and a linear gradient elution of 25 to 55% acetonitrile in 40 min. Collected fractions were lyophilized to obtain the objective compounds.

Synthesis of TTet59–A. Solid–phase synthesis was performed with 42.0 mg Fmoc–β–Wang resin (0.60 mmol/ g) and finally **TTet59–A** (3.0 mg, 6.8 x 10^{-4} mmol, 2.7% yield) was obtained as a pale yellow solid. Analytical HPLC: t_R = 16.9 min. ESI–TOF–MS *m/z* calcd for $C_{200}H_{246}N_{79}O_{42}^{5+}$ $[M + 5H]^{5+}$ 885.1903 found 885.1820.

Synthesis of TTet59–B. Solid–phase synthesis was performed with 43.0 mg Fmoc–β–Wang resin (0.60 mmol/ g) as a pale yellow solid. The resulting crude polyamide was purified as described to afford **TTet59–B** (8.9 mg, 2.0 x 10^{-3} mmol, 7.7% yield). Analytical HPLC: t_R = 17.1 min. ESI–TOF–MS *m/z* calcd for $C_{203}H_{252}N_{79}O_{42}^{5+}$ $[M + 5H]^{5+}$ 893.5997 found 893.5936.

Synthesis of TAMRA TTet59–A. TTet59–A (0.9 mg) and 5–carboxytetramethyl-rhodamine succinimidyl ester (0.2 mg) were dissolved in DMF (250 μL) and DIEA (0.40 μL), followed by mixing at room temperature with shielding the light. After checking this reaction had been finished, the reaction mixture was purified by reversed–phase HPLC in 0.1% TFA in water with acetonitrile as the eluent at a flow rate of 1.0 mL/min and a linear gradient elution of 30 to 70% acetonitrile in 30 min, followed by lyophilization of the collected fractions to afford **TAMRA TTet59–A** (1.1 mg) as a purple powder. Analytical HPLC: t_R = 18.2 min. ESI–TOF–MS *m/z* calcd for $C_{225}H_{267}N_{81}O_{46}^{6+}$ $[M + 6H]^{6+}$ 806.5169 found 806.5205.

Synthesis of TAMRA TTet59–B. TTet59–B (2.6 mg) and 5–carboxytetramethyl-rhodamine succinimidyl ester (0.9 mg) were dissolved in DMF (272 μL) and DIEA (0.60 μL). The reaction and purification were performed as described above to afford **TAMRA TTet59–B** (2.0 mg) as a purple powder. Analytical HPLC: t_R = 18.3 min. ESI–TOF–MS *m/z* calcd for $C_{228}H_{274}N_{81}O_{46}^{7+}$ $[M + 7H]^{7+}$ 697.4508 found 697.4565.

Synthesis of Biotin TDi59–A. TDi59–A (1.9 mg) and EZ–Link NHS–PEG_{12}–Biotin 0.9 mg were dissolved in DMF (159 μL) and DIEA (0.30 μL). The reaction and purification were performed as described above to afford **Biotin TDi59–A** (0.4 mg) as a white solid. Analytical HPLC: t_R = 17.4 min. ESI–TOF–MS *m/z* calcd for $C_{139}H_{198}N_{44}O_{35}S^{4+}$ $[M + 4H]^{4+}$ 768.8691 found 768.8644.

Synthesis of Biotin TTet59–B. TTet59–B (0.8 mg) and EZ–Link NHS–PEG_{12}–Biotin 0.1 mg were dissolved in DMF (101 μL) and DIEA (0.20 μL). The reaction and purification were performed as described above to afford **Biotin TTet59-B** (0.5 mg) as

a white solid. Analytical HPLC: t_R = 19.0 min. ESI–TOF–MS *m/z* calcd for $C_{240}H_{321}N_{82}O_{57}S^{7+}$ $[M + 7H]^{7+}$ 756.4918 found 756.4987.

SPR Assays. The SPR assays were performed as described in previous reports.[1b,c] Biotinylated DNA was immobilized to streptavidin–coated sensor chip at a flow rate of 10 μL/min to obtain the desired immobilization level (up to approximately 1000 RU rise). The assays were performed with HBS–EP (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20) containing 0.1% DMSO at 25 °C. A series of sample solutions with various concentrations were prepared in HBS–EP buffer containing 0.1% DMSO and injected at a flow rate of 20 μL/min. To calculate association rates (k_a), dissociation rates (k_d), and dissociation constants (K_D), data processing was performed with 1:1 binding with mass transfer model using BIAevaluation 4.1 program.

Telomeres Staining and Calculation of S/N Ratios

HeLa 1.3 Cell Spreads Treated with Fluorescent Py–Im polyamide Probes.[1,3] HeLa 1.3 cells were maintained in DMEM medium containing 10% fetal bovine serum (FBS) at 37 °C (5% CO_2). HeLa 1.3 cells were blocked mitotically by adding 0.2 μg/mL colcemid and incubated overnight. The cells were then swollen by treatment with hypotonic buffer (0.075 M KCl) for 15 min at room temperature, fixed with methanol/acetic acid (3:1) solution for 5 min. After centrifugation, cell pellet was again suspended in the methanol/acetic acid (3:1) solution. The cell suspension was spread on coverslips and air–dried at room temperature for 5 min and then at 60 °C for 30 min. The cell spread coverslips were kept at 4 °C until use.

The dried spread coverslips were soaked in HEN buffer (10 mM HEPES pH 7.5, 1 mM EDTA and 100 mM NaCl) over night at 4 °C before use. For blocking, the spread coverslips were treated with 10% NGS in TE buffer 1 (10 mM Tris–HCl pH 7.5, 1 mM EDTA) for 30 min at room temperature. After brief washing with TE buffer 1, the spread coverslips were incubated with 10% NGS, 75 nM fluorescent Py–Im polyamide probe and 0.5 μg/ml DAPI in TE buffer 1 containing 10% DMF for 1 h at 37 °C. After washing with TEN100 (10 mM Tris–HCl pH 7.5, 1 mM EDTA and 100 mM NaCl) buffer (five times for 3 min), the spread coverslips were mounted in PPDI solution (10 mM HEPES pH 7.5, 100 mM KCl, 1 mM $MgCl_2$, 80% glycerol, 1 mg/mL paraphenylene diamine) and the coverslips were sealed with a nailporish. Sectioning images were recorded with DeltaVision and projected (“Quick Projection” tool) without deconvolution to obtain telomeric signals and background signals. For quantification of the telomere signals in Figure 2, the telomere signals yielded by the polyamides were

extracted based on threshold values using the Softworks software (Applied Precision). The maximum intensity values of signals in the extracted telomere regions were then used as telomere signals. For the background signals, ROIs which were same shape of each telomere were set next to the defined each telomere signals in the chromosomes, and maximum values of the signals were used as background signals. Box plot and dot plot were created in R software.

Formaldehyde–fixed HeLa 1.3 Cells Treated with Fluorescent Py–Im Polyamide Probes.[1] For polyamide staining, the HeLa 1.3 cells were grown on coverslips coated with polylysine. The cell coverslips were washed in phosphate–buffered saline (PBS) twice and fixed with 1.85% formaldehyde in PBS for 15 min at room temperature. The fixed cell coverslips were then treated with 50 mM glycine in PBS for 5 min, and permeabilized with 0.5% Triton X–100 in PBS for 5 min. After briefly washing with HMK buffer (10 mM HEPES pH 7.5, 1 mM $MgCl_2$, 100 mM KCl) twice, the coverslips were soaked in HEN buffer for overnight and kept at 4 °C until use. For blocking, the cell coverslips were treated with 10% NGS in TE buffer 1 for 30 min at room temperature. After brief washing with TE buffer 1, the cell coverslips were incubated with 10% NGS, 15 nM fluorescent Py–Im polyamide probe and 0.5 μg/ml DAPI in TE buffer 1 containing 10% DMF for 1 h at 37 °C. After washing with TEN200 buffer (10 mM Tris–HCl pH 7.5, 1 mM EDTA, and 200 mM NaCl) (five times for 3 min), the mounting and subsequent image acquisitions were performed as described above. For quantification of the telomere signals in Figure 3, they were extracted from the images as described previously.[1] The telomere signals yielded by the polyamides were extracted based on threshold values using the Softworks software (Applied Precision). The maximum intensity values of signals in the extracted telomere regions were then used as telomere signals. For the background signals, 10 squares (10 pixels x 10 pixels) were randomly set outside the defined telomere signals and mean values of the signals in the squares were used as background signals. Box plot and dot plot were created in R software.

Identification of TDi59-A and TTet59-B Binding Sites in Genome

Pull–down and Affinity Purification of DNA bound by Biotin TDi59–A and Biotin TTet59–B.[4] Human foreskin BJ fibroblast cells were maintained in 8 mL DMEM medium containing 10% FBS, 1x non–essential amino acids, penicillin (100 units/mL) and streptomycin (100 μg/mL) at 37 °C (5% CO_2). 75−80% confluent cells (2 x 10^6 cells) were washed with PBS and isolated by tripsinization for 3 min. 3 plates of 2 x 10^6

cells were used. Each isolated 2 x 10^6 cells were again washed with cold PBS and the cell pellet was suspended in 5 mL ice cold NP−40 lysis buffer (10 mM Tris–HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.5% Nonidet P–40, 0.15 mM spermine, 0.5 mM spermidine and 0.015x protease inhibitor cocktail) and incubated on ice for 5 min. The nuclei were pelleted by centrifugation at 300 g for 10 min and then carefully dissolved in 900 μL binding buffer (10 mM Tris–HCl pH 8.0, 5 mM $MgCl_2$, 1 mM Dithiothreitol, 0.3 M KCl, 0.3x protease inhibitor cocktail and 10% glycerol)[4b] and incubated at 4 °C for 1 h. 100 μL of the binding buffer containing 1 μL of 400 μM biotinylated Py–Im polyamide probe dissolved in DMSO was added to the extracted nuclei solution. Each 2 x 10^6 cells nuclei were incubated with biotinylated Py–Im polyamide probe (400 nM final concentration) dissolved in DMSO (0.1% final concentration) at 4 °C for 16 h. After incubation, the nuclei were pelleted by centrifugation at 5200 rpm for 10 min and then washed with 1 mL washing buffer A (10 mM Tris–HCl pH 7.4, 15 mM NaCl, 60 mM KCl, 0.15 mM spermine, 0.5 mM spermidine and 0.1x protease inhibitor cocktail). After removal of washing buffer A by centrifuge at 5200 rpm for 10 min, the pellets were resuspended in 50 μL MNase buffer (20 mM Tris–HCl pH 8.0, 5 mM NaCl, 2.5 mM $CaCl_2$). Then 4 units micrococcal nuclease (MNase) in 50 μL MNase buffer containing 43 μg RNase A and 0.2x protease inhibitor cocktail was added to the solution, followed by incubation at 37 °C for 30 min to digest to mononucleus.[4b] After this digestion 2 μL of 0.5 M EDTA was added to quench MNase and then histone protein was removed by addition of 4 μL of 20 mg/mL proteinase K solution and incubation at 37 °C for 30 min. After MNase digestion and proteinase K treatment, the suspension was mixed with equal volume of modified COSMIC[4a] buffer (20 mM Tris–HCl pH 8.1, 2 mM EDTA, 150 mM NaCl, 0.3x protease inhibitor cocktail, 1% Triton–X 100, and 0.1% SDS) containing 0.50 mg of suspended streptavidin−coated magnetic beads (C1 beads) and then incubated at 4 °C for 20 h to pull down target DNA. Preparation of the beads was performed as follows: after removing the suspension solution from Dynabeads® MyOne™ Streptavidin C1, they were washed with modified COSMIC buffer and resuspended in it.

After the pull down, beads–bound DNA was isolated by reported affinity purification.[4] Beads were washed once with 500 μL of washing buffer 1 (10 mM Tris–HCl pH 8.0, 1 mM EDTA, 0.3% SDS) for 7 min, once with 500 μL of washing buffer 2 (10 mM Tris–HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% Nonidet P–40) for 5 min, twice with 500 μL washing buffer 3 (10 mM Tris–HCl pH 7.5, 1 mM EDTA, 0.1% Nonidet P–40) for 5 min and twice with 500 μL TE buffer 2 (10 mM Tris–HCl pH 8.0 and 1 mM EDTA) for 5 min. Beads were resuspended in elution buffer 1 (10 mM Tris–

HCl pH 7.6, 0.4 mM EDTA and 100 mM KOH) and DNA was eluted from magnetic beads after heating at 90 °C for 30 min. The remaining DNA with the beads was eluted using elution buffer 2 (2% SDS, 100 mM $NaHCO_3$ and 3 mM biotin) with the heating at 65 ^oC for 8 h. The detached DNA was purified with QIAquick PCR Purification Kit following the manufacture protocol and then quantified with a Nanodrop. Isolated DNA were pooled to get optimum DNA concentration for the sequencing library construction. Biological replicates were assessed.

High–throughput Sequencing.[4b] Sequencing libraries of the collected DNA were prepared with standard Ion Xpress™ Plus gDNA Fragment Library Preparation reagents (Thermo Fisher Scientific Inc.). A1 and P1 adapter–ligated enriched DNA was amplified and purified, followed by quality and quantity check with Agilent 2100 Bioanalyzer. The qualified libraries were used for high–throughput sequencing. Following the manufacture's protocol, the sequencing was performed, starting with template preparation using Ion PI™ template OT2 200 kit in Ion one touch2 system (Thermo Fisher Scientific Inc.). The templates were then enriched using Ion one touch ES. The enriched libraries were sequenced with 300 flow of single read by Ion Proton™ Sequencer using Ion PI™ Sequencing 200 kit v3/ Ion PI chip by following the manufacturer's guidelines. 25–30 million post filtered reads per library were produced. The data were handled by employing standard program packages in the Ion torrent suit. Torrent Mapping Alignment Program 4.4.2 (TMAP) was used for aligning reads with the hg38 reference genome, and enriched peaks was called using MACS 1.4.2.[5a] The enriched genomic regions were visualized using UCSC genome browser (http://genome.ucsc.edu/). Homer motif analysis program was used for the *de novo* motif analysis.[5b]

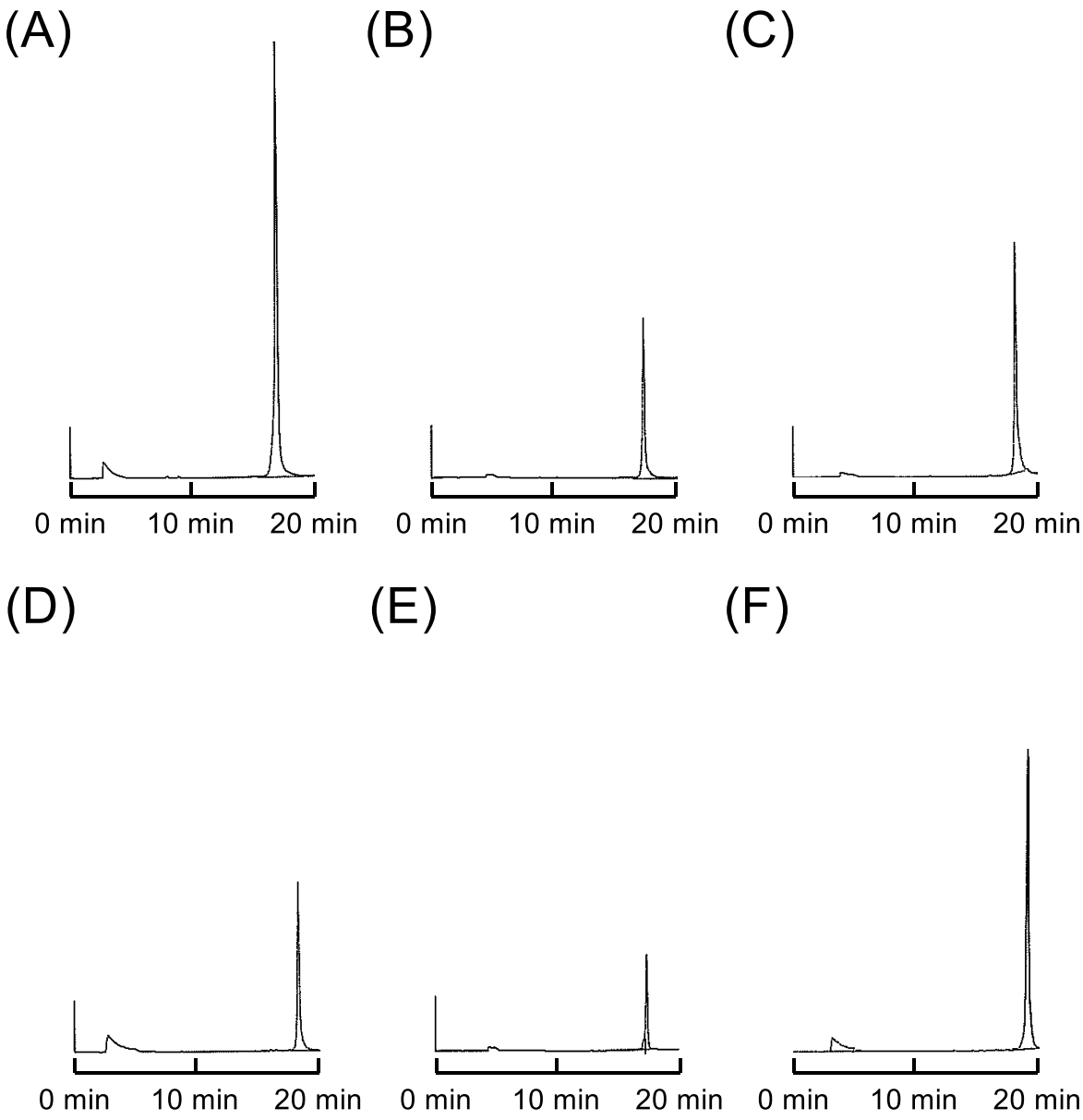


Figure S1. Analytical HPLC profiles of Py–Im polyamides (A: **TTet59**–**A**, B: **TTet59**–**B**, C: **TAMRA TTet59**–**A**, D: **TAMRA TTet59**–**B**, E: **Biotin TDi59**–**A**, F: **Biotin TTet59**–**B**).

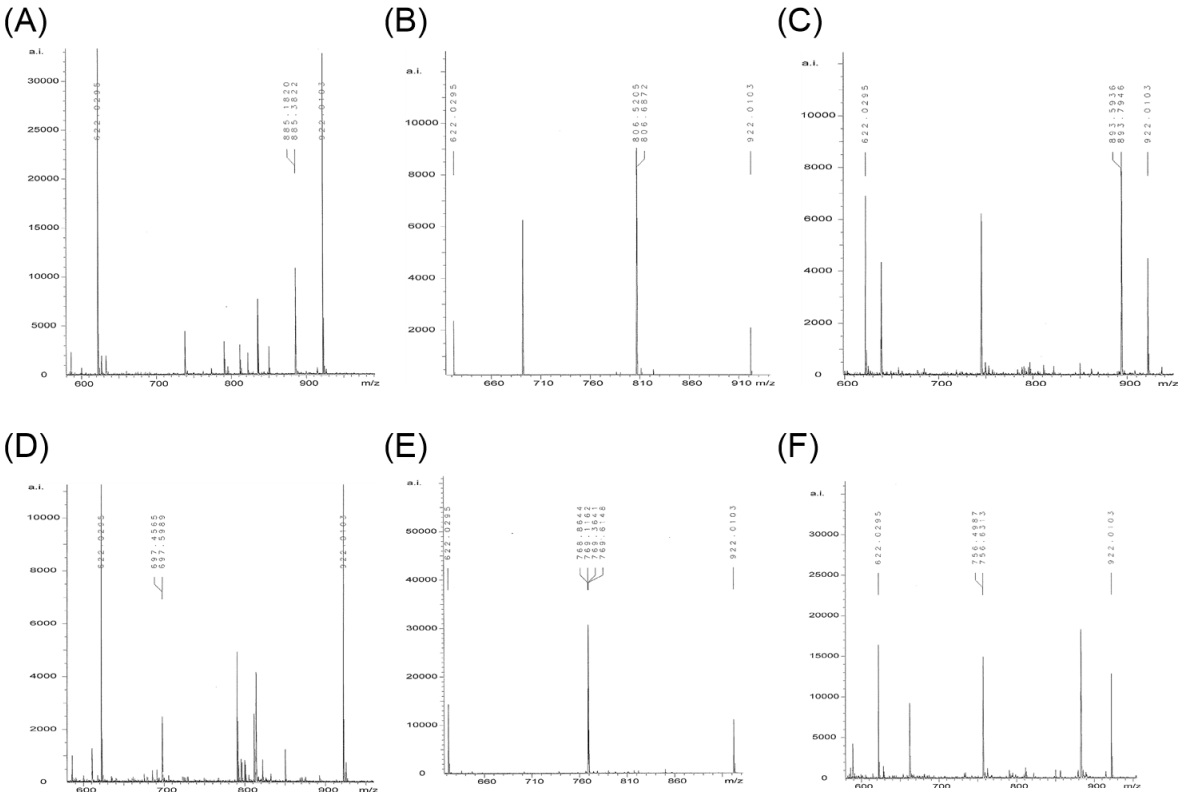


Figure S2. ESI–TOF MS spectra of Py–Im polyamides (A: **TTet59**–**A**, B: **TTet59**–**B**, C: **TAMRA TTet59**–**A**, D: **TAMRA TTet59**–**B**, E: **Biotin TDi59**–**A**, F: **Biotin TTet59**–**B**).

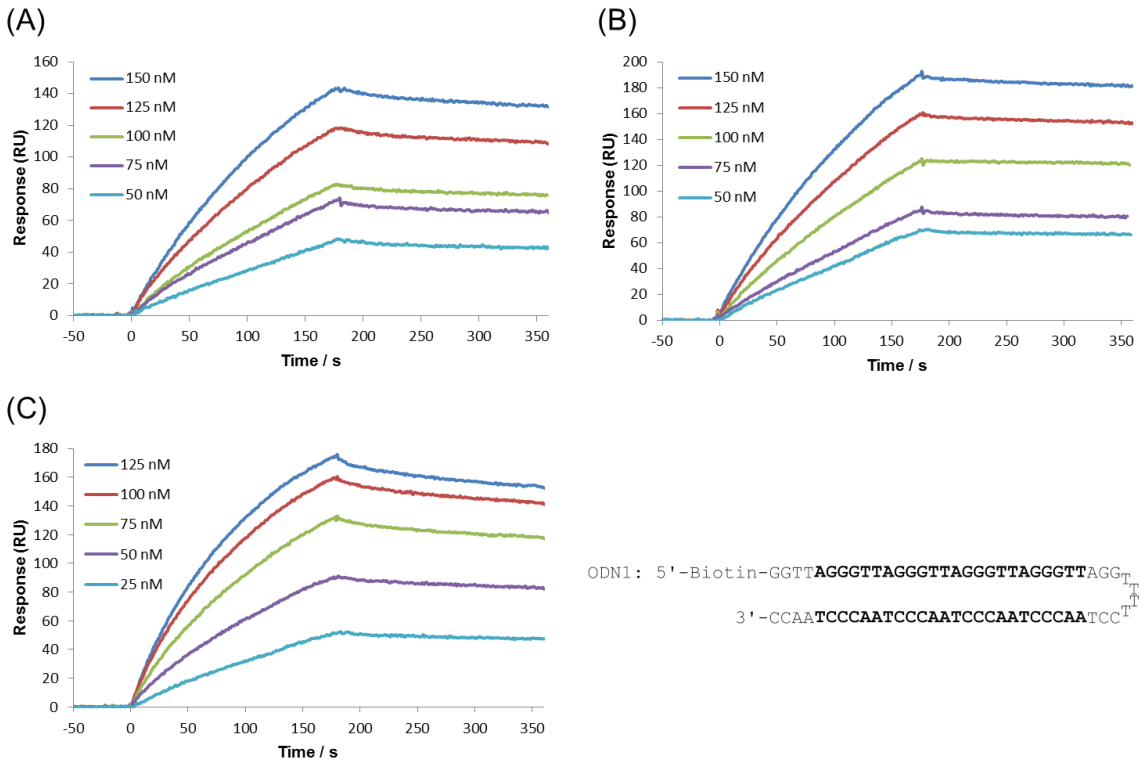


Figure S3. SPR sensorgrmas for the interaction between ODN-1 and (A) **TTet59**–**A**, (B) **TTet59**–**B** or (C) **TTri59**–**A**.

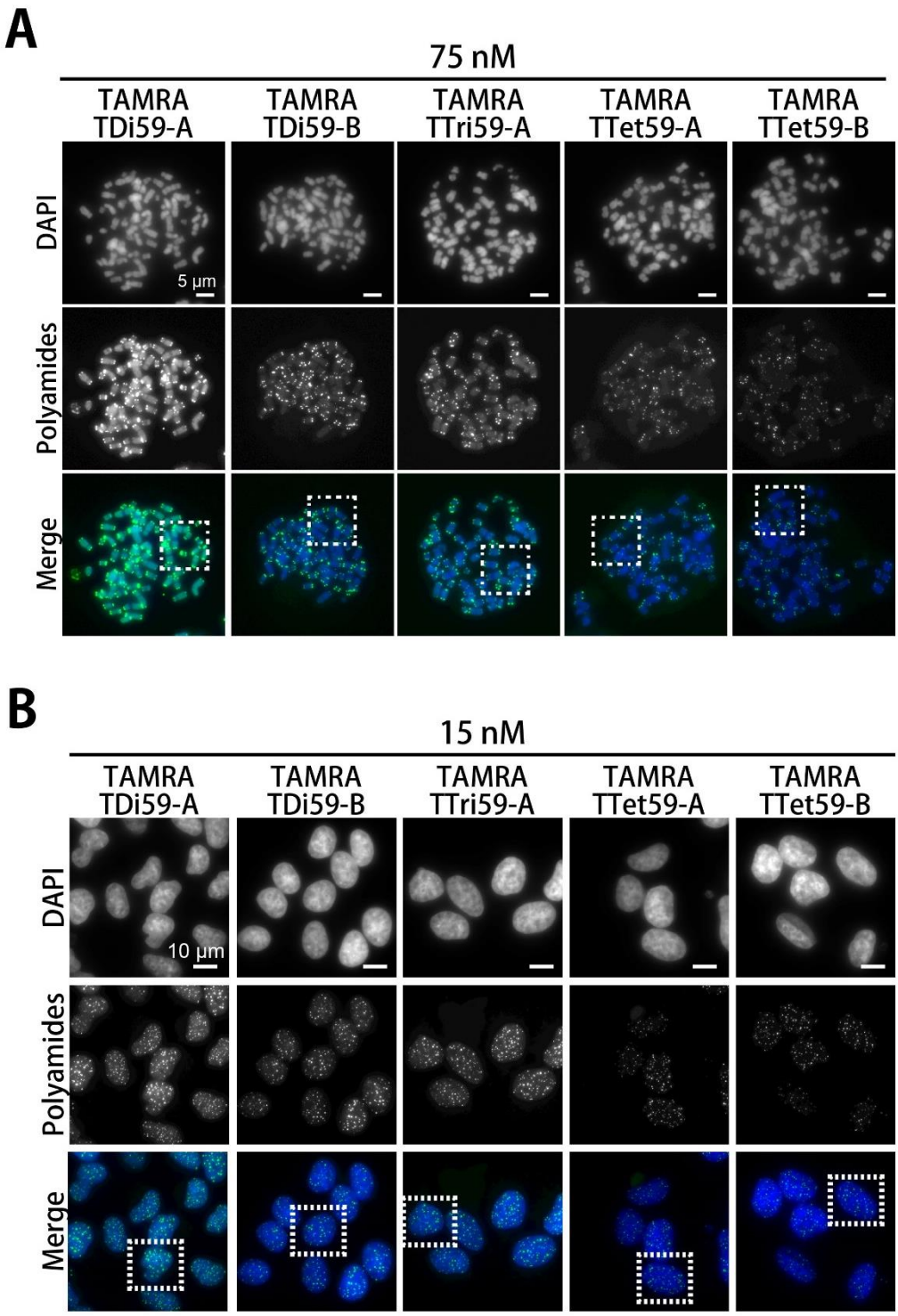


Figure S4. Full images of HeLa 1.3 cell spreads in Figure 2 and HeLa 1.3 cells in Figure 3. Enlarged images of the boxed regions in (A) and (B) are shown in Figure 2 and 3, respectively.

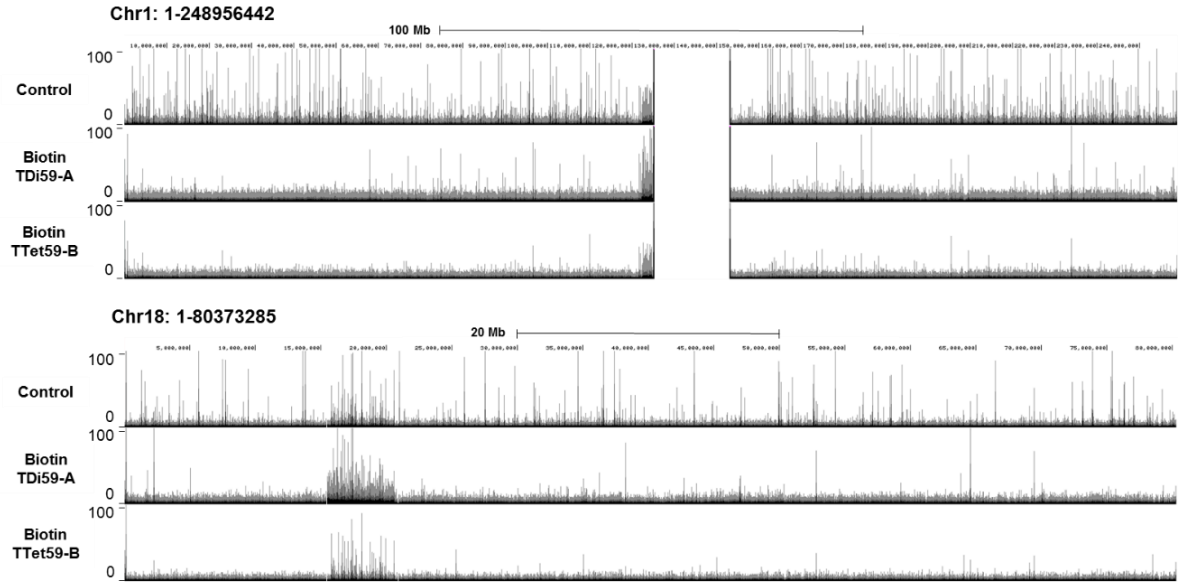
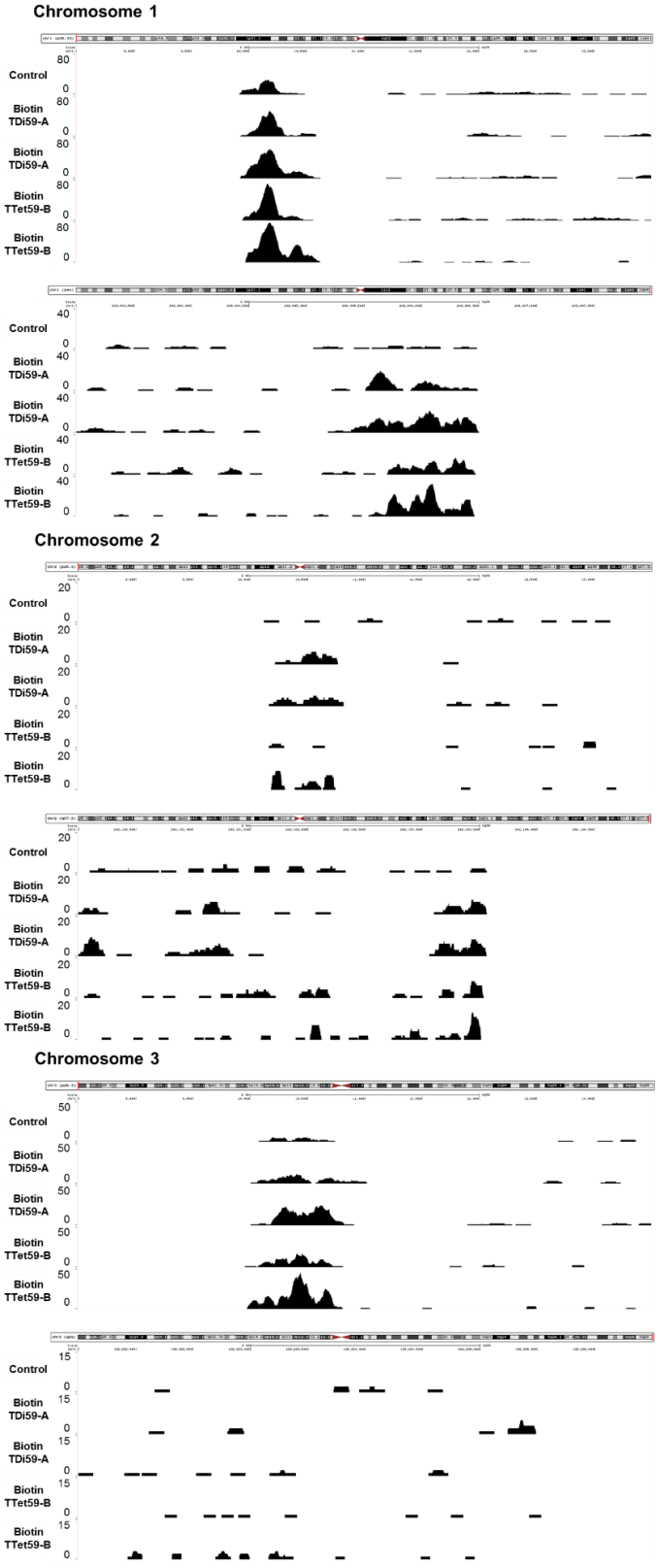
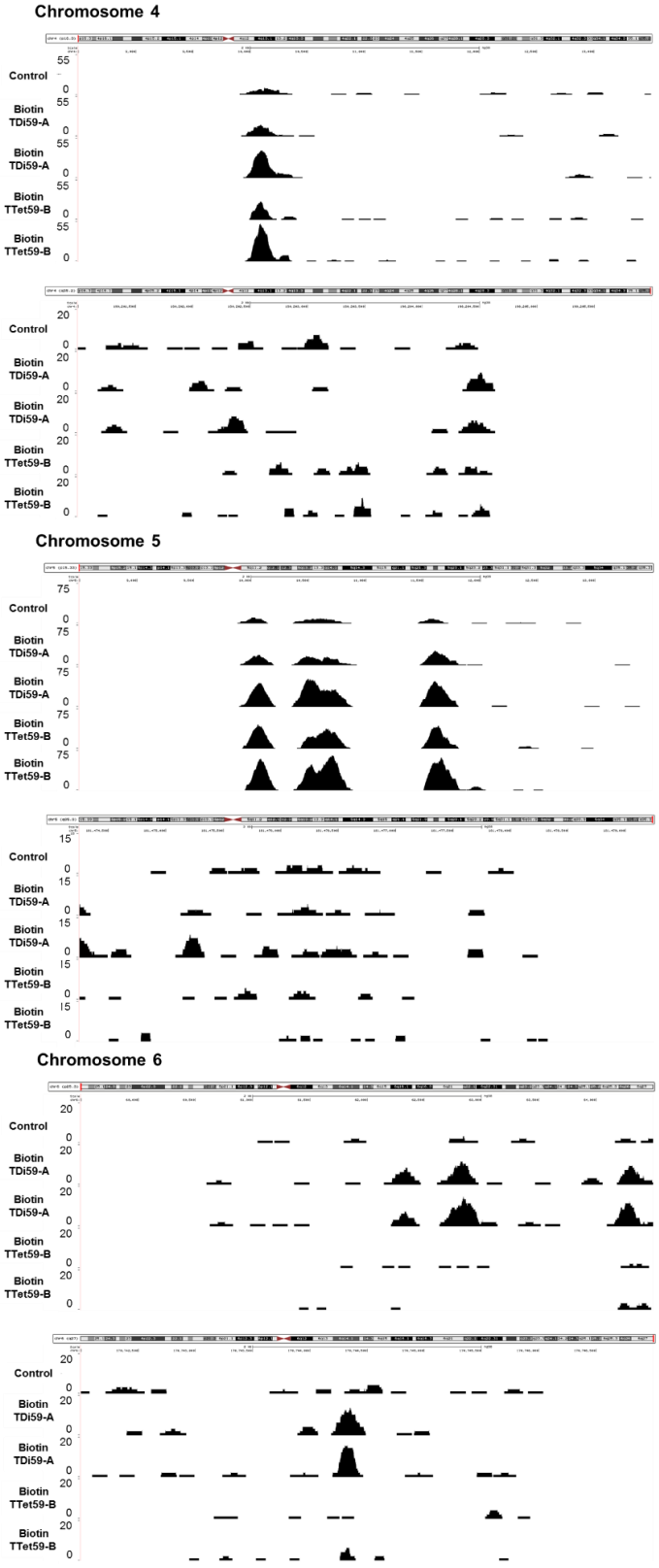
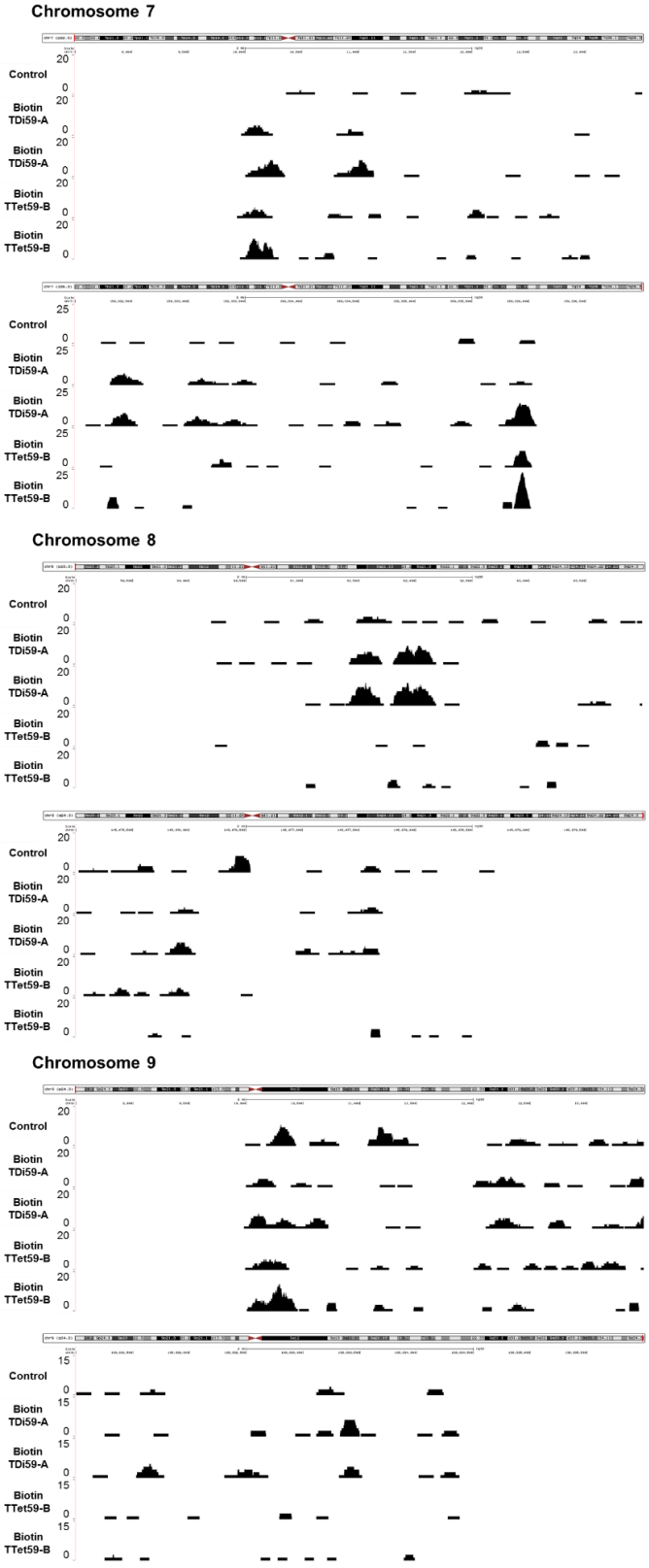
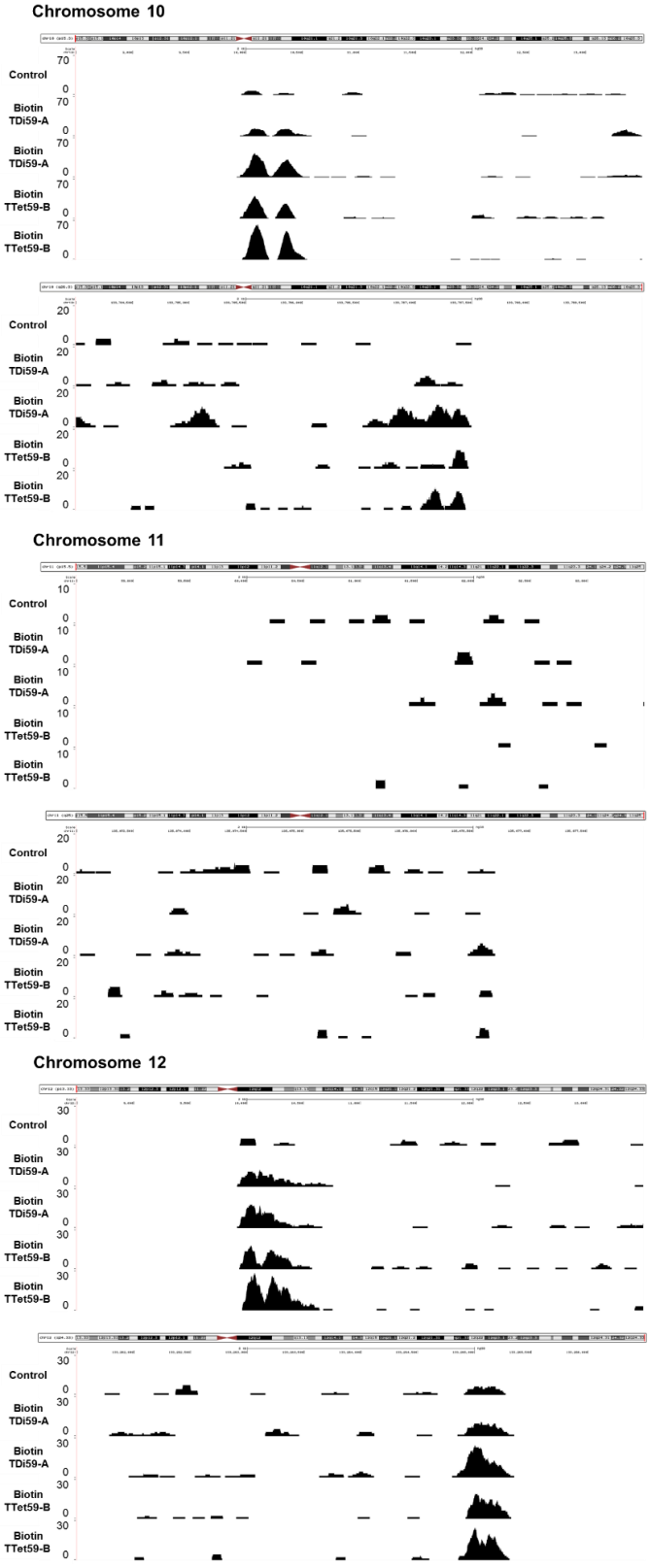


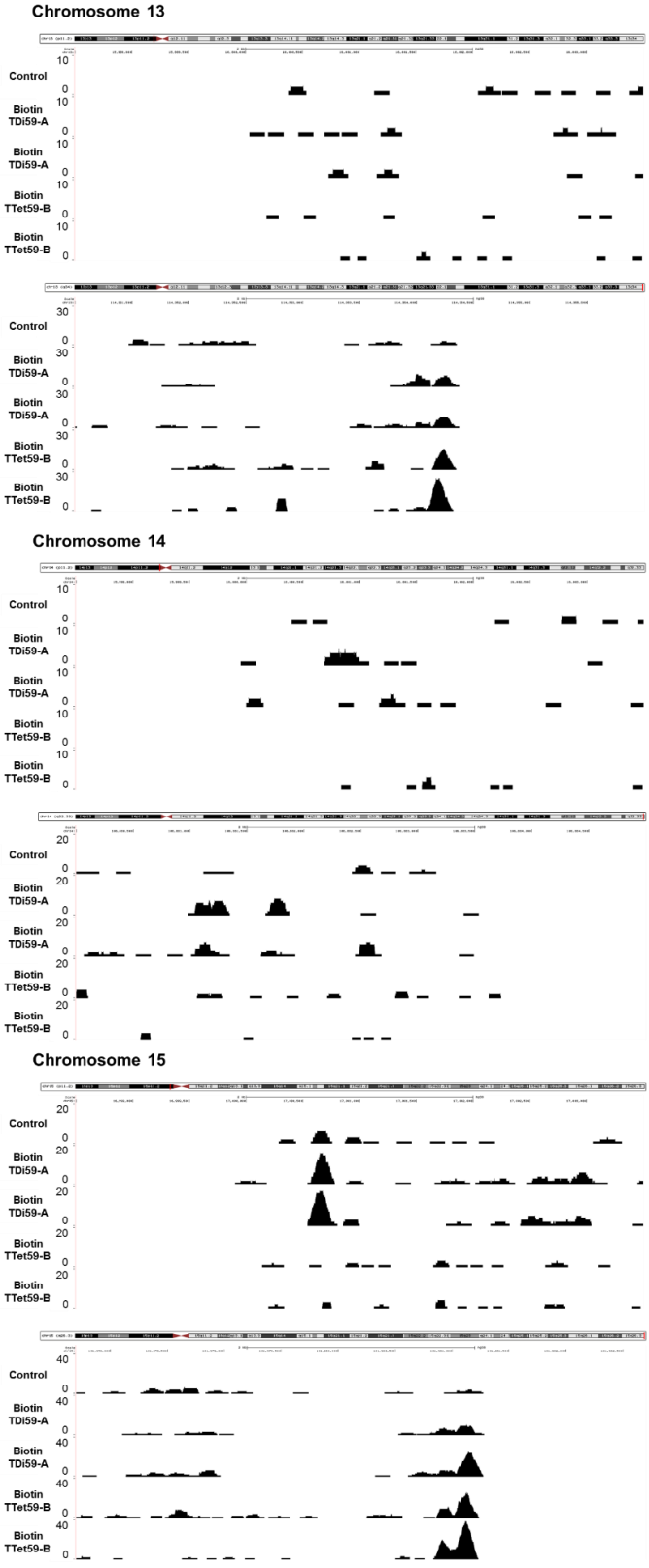
Figure S5. Mapping of the biotinylated Py–Im polyamide probes binding and enriched sites in the whole of chromosomes 1 and 18.

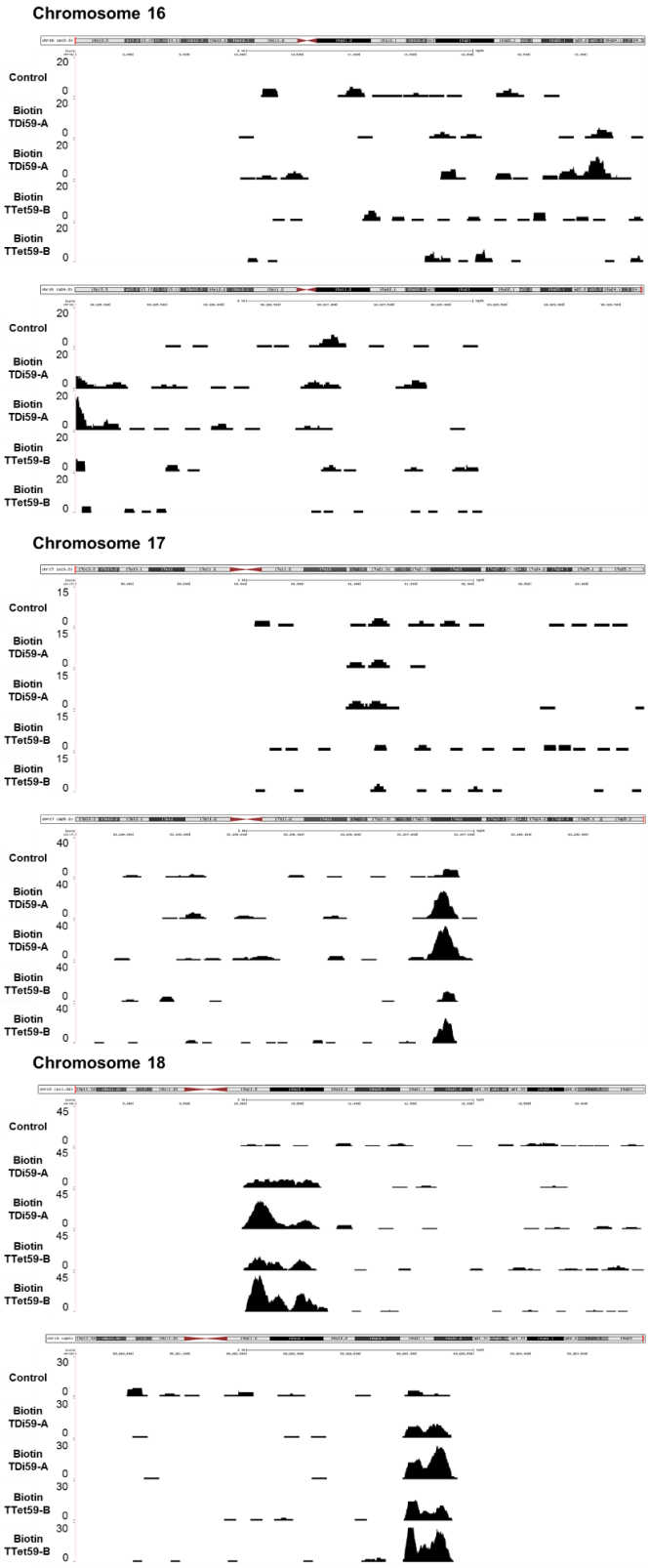


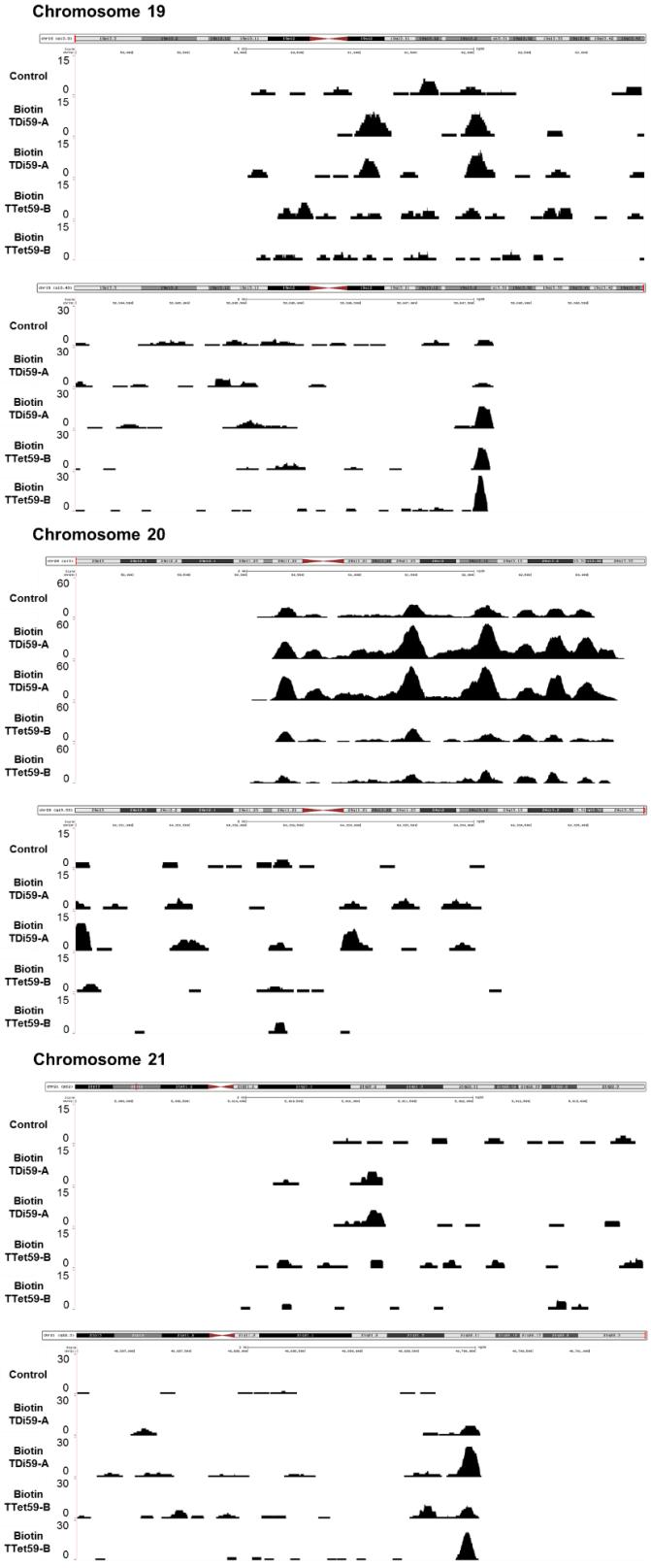












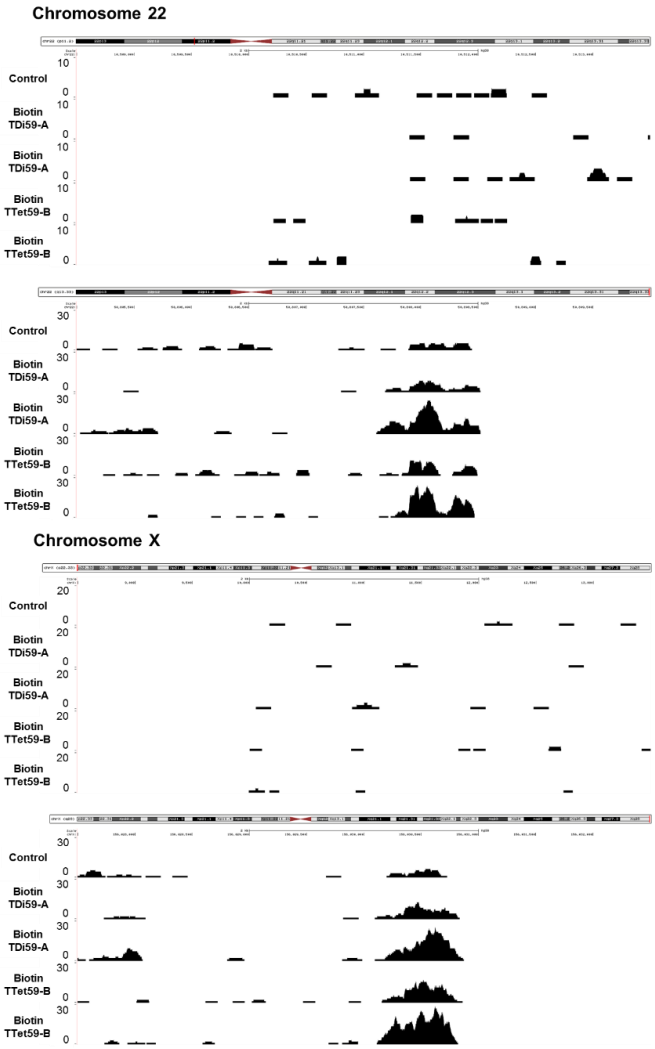


Figure S6. The enrichment of **Biotin TDi59–A** and **Biotin TTet59–B** at both termini for all chromosomes.

References

(1) (a) Kawamoto, Y.; Bando, T.; Kamada, F.; Li, Y.; Hashiya, K.; Maeshima, K.; Sugiyama, H. *J. Am. Chem. Soc.* **2013**, *135*, 16468–16477. (b) Hirata, A.; Nokihara, K.; Kawamoto, Y.; Bando, T.; Sasaki, A.; Ide, S.; Maeshima, K.; Kasama, T.; Sugiyama, H. *J. Am. Chem. Soc.* **2014**, *136*, 11546–11554. (c) Kawamoto, Y.; Sasaki, A.; Hashiya, K.; Ide, S; Bando, T.; Maeshima, K; Sugiyama. H. *Chem. Sci.* **2015**, *6*, 2307–2312. (d) Sasaki, A.; Ide, S.; Kawamoto, Y.; Bando, T.; Murata, Y.; Shimura, M.; Yamada, K.; Hirata, A.; Nokihara, K.; Hirata, T.; Sugiyama, H.; Maeshima, K. *Sci. Rep.* **2016**, *6*, 29261. (e) Maeshima, K.; Janssen, S.; Laemmli, U. K. *EMBO J.* **2001**, *20*, 3218−3228.
(2) Minoshima, M.; Bando, T.; Sasaki, S.; Fujimoto, J.; Sugiyama, H. *Nucleic Acids Res.* **2008**, *36*, 2889–2894.
(3) (a) Maeshima, K.; Laemmli, U. K. *Dev. Cell* **2003**, *4*, 467–480. (b) Maeshima, K.; Yahata, K.; Sasaki, Y.; Nakatomi, R.; Tachibana, T.; Hashikawa, T.; Imamoto, F.; Imamoto, N. *J. Cell Sci.* **2006**, *119*, 4442–4451.
(4) (a) Erwin, G. S.; Bhimsaria, D.; Eguchi, A.; Ansari, A. Z. *Angew. Chem. Int. Ed.* **2014**, *53*, 10124−10128. (b) Chandran, A.; Syed, J.; Taylor, R. D.; Kashiwazaki, G.; Sato, S.; Hashiya, K.; Bando, T.; Sugiyama, H. *Nucleic Acids Res.* **2016**, *44*, 4014–4024. (c) Chandran, A.; Syed, J.; Li, Y.; Sato, S.; Bando, T.; Sugiyama, H. *ChemBioChem.* **2016**, DOI: 10.1002/cbic.201600274.
(5) (a) Feng, J.; Liu, T.; Qin, B.; Zhang, Y.; Liu, X. S. *Nat. Protoc.* **2012**, *7*, 1728–1740. (b) Heinz, S.; Benner, C.; Spann, N.; Bertolino, E.; Lin, Y. C.; Laslo, P.; Cheng, J. X.; Murre, C.; Singh, H.; Glass, C. K. *Mol. Cell* **2010**, *38*, 576–589.